Pragmatica dell'Intenzione Distribuita

Atti linguistici senza soggetto unitario

Il soggetto silenzioso: anatomia di un presupposto

I. L'assioma nascosto

Ogni teoria ha i propri assiomi manifesti e i propri presupposti opachi, quelli che non vengono mai messi sul tavolo perché sembrano troppo ovvi per meritare enunciazione. La pragmatica del linguaggio ordinario — quella che prende forma con Austin a Oxford nel dopoguerra e si consolida attraverso Grice e Searle fino a diventare l'ossatura concettuale di tanta filosofia del linguaggio, linguistica e teoria della comunicazione — custodisce nel suo nucleo un presupposto di questo genere, mai esplicitato con sufficiente rigore eppure pervasivo in ogni piega dell'argomentazione: che il parlante sia un soggetto unitario, psicologicamente integrato, titolare di intenzioni che precedono e causano la sua attività verbale.

Questo presupposto non è innocuo. È, piuttosto, la condizione di possibilità dell'intera architettura teorica. Toglietelo, e l'edificio non crolla subito — i mattoni rimangono al loro posto — ma le fondamenta vacillano, e con esse vacilla una certa idea di cosa significhi comunicare, di come nasca l'obbligazione verbale, di dove risieda la responsabilità di un discorso.

Chiamiamo questo presupposto il *principio del soggetto unitario* (PSU): l'assunzione che ogni atto linguistico sia riconducibile a un agente singolare la cui struttura psicologica interna è, per quanto complessa, sufficientemente coerente da consentire l'attribuzione di intenzioni

discrete, coscienti e causalmente efficaci rispetto al comportamento verbale. Il PSU non è mai stato articolato come tale dai fondatori della pragmatica. La sua formulazione esplicita richiede già uno sguardo critico che i fondatori stessi non esercitarono — non perché fossero superficiali, ma perché il PSU era così profondamente radicato nell'Zeitgeist filosofico da sembrare non un'ipotesi empirica ma una verità concettuale.

La presente monografia intende esaminare le conseguenze di un esperimento mentale radicale: cosa accade alla pragmatica se il PSU viene rimosso non per arbitrio critico ma per ragioni teoriche sostanziali? E cosa emerge, al suo posto, se si tenta di ricostruire una teoria degli atti linguistici adeguata a descrivere agenti genuinamente distribuiti — organizzazioni, istituzioni, comunità, reti neurali biologiche e artificiali — per i quali non esiste un centro intenzionale univoco?

II. Austin e la performatività: forza senza fonte

John Langshaw Austin introduce la distinzione tra enunciati constativi e performativi in *How to Do Things with Words*, raccogliendo le William James Lectures del 1955. L'intuizione centrale è straordinariamente feconda: esistono enunciati che non descrivono stati di cose ma li producono, non riferiscono azioni ma le compiono. "Prometto di venire", "Dichiaro aperta la seduta", "Ti nomino cavaliere" non sono asserzioni verificabili come vere o false ma *atti* il cui compimento coincide con la loro enunciazione in condizioni appropriate.

La teoria delle *condizioni di felicità* che Austin articola per dar conto di quando un atto performativo riesce o fallisce è una delle costruzioni filosofiche più eleganti del Novecento. Un performativo può mancare il suo scopo in modi diversi: per vizio di procedura, per insincerità del parlante, per inosservanza delle convenzioni costitutive. Ed è proprio qui, nella categorizzazione dei fallimenti, che il PSU fa la sua prima apparizione implicita ma decisiva.

Quando Austin distingue tra atti che *misfanno* (misfire) e atti che costituiscono *abusi* (abuse), introduce una differenza che presuppone l'integrità psicologica del parlante. Un abuso — diciamo, una promessa formulata senza la reale intenzione di mantenerla — è tale perché il parlante ha un'intenzione nascosta che diverge dall'intenzione ostentata. Questa struttura duale richiede logicamente un soggetto capace di

ospitare contemporaneamente stati intenzionali in conflitto: l'intenzione performativa di produrre un vincolo e l'intenzione reale di non onorarlo. Ma richiede anche, implicitamente, che questi stati siano attribuibili a un *unico* portatore psicologico, pena il dissolvimento della categoria stessa di abuso.

Si consideri ora un caso elementare di distribuzione: un consiglio di amministrazione approva all'unanimità una delibera. La delibera ha la forma e la forza di una promessa pubblica verso gli azionisti. Ma nessuno dei singoli consiglieri la *voleva* esattamente così: alcuni avrebbero preferito un impegno più cauto, altri più ambizioso, alcuni si sono astenuti mentalmente pur votando a favore per ragioni tattiche. Dove risiede l'intenzione di questa promessa? Chi è il soggetto dell'eventuale abuso se l'impegno non verrà rispettato? Austin non aveva strumenti per rispondere perché non aveva nemmeno gli strumenti per formulare la domanda.

La transizione che Austin opera dalla dicotomia constativo/performativo alla teoria generale degli atti illocutori nella seconda metà del libro acuisce piuttosto che risolvere il problema. Ogni enunciato ha una *forza illocutoria* — è un'asserzione, una domanda, un comando, una promessa, un'avvertenza — e questa forza è determinata in parte dalla convenzione linguistica e in parte dall'intenzione del parlante. La locuzione convenzionale è condizione necessaria ma non sufficiente: ci vuole anche un agente che la *usi* con una determinata intenzione illocutoria. Il soggetto intenzionale rimane insostituibile nell'economia della teoria.

Eppure Austin lascia aperta, quasi involontariamente, una crepa. Quando discute il modo in cui la forza illocutoria viene *compresa* dall'interlocutore, ammette che la convenzione gioca un ruolo primario rispetto all'intenzione individuale: certe formule hanno una forza illocutoria convenzionalmente stabilita che il parlante non può modificare a piacere con una semplice ridescrizione intenzionale. Un giudice che pronuncia "dichiaro l'imputato colpevole" non può poi affermare di aver inteso fare semplicemente un'asserzione ipotetica: la forza dell'atto è determinata dal contesto istituzionale, non dall'intenzione psicologica individuale. Questa osservazione, che Austin non sviluppa in modo sistematico, contiene in nuce la possibilità di una pragmatica post-intenzionalista.

III. Grice e l'intenzione comunicativa: un circolo virtuoso che esclude il distribuito

Se Austin lascia la crepa aperta, Paul Grice la richiude con precisione tecnica. La sua teoria dell'intenzione comunicativa non-naturale (meaning-nn) è il tentativo più rigoroso di fondare il significato linguistico sull'intenzionalità del parlante, e proprio per questo costituisce il punto più vulnerabile della tradizione classica di fronte alla sfida del distribuito.

Il celebre schema griceano del significato-nn afferma che un parlante S intende non-naturalmente comunicare qualcosa con un'emissione U se e solo se: S intende produrre una risposta r in un uditore A; S intende che A riconosca questa intenzione; S intende che il riconoscimento di questa intenzione da parte di A costituisca (almeno in parte) la ragione per cui A produce r. La struttura è iterativa e riflessiva: l'intenzione comunicativa è un'intenzione a essere riconosciuta come tale, un'intenzione la cui efficacia causale passa necessariamente attraverso la sua trasparenza all'interlocutore.

Questa costruzione è di straordinaria acutezza filosofica. Ma richiede un presupposto che non viene mai reso esplicito: che il soggetto S sia capace di ospitare simultaneamente intenzioni di diversi ordini (intenzione di produrre r; intenzione che A riconosca questa intenzione; intenzione che questo riconoscimento causi r) e di mantenerle in una relazione strutturata e coerente. Questo è possibile solo se S è un sistema psicologico sufficientemente integrato da poter gerarchizzare e coordinare stati intenzionali complessi. Un'organizzazione, un'istituzione, una rete di agenti non possiede questa capacità in modo ovvio. Dove risiede, in un consorzio universitario che emette un comunicato stampa, la struttura iterativa di intenzioni di Grice?

La risposta standard — che l'organizzazione agisce attraverso i suoi rappresentanti, e che le intenzioni dei rappresentanti sono le intenzioni dell'organizzazione — è troppo rapida. Lascia irrisolto il problema di come le intenzioni multiple, divergenti e spesso contrastanti dei diversi agenti che concorrono alla produzione di un atto linguistico istituzionale si *integrino* in un'unica intenzione comunicativa attribuibile all'ente. La risposta giuridica — che l'ente risponde attraverso i suoi organi statutariamente competenti — è una soluzione normativa, non una spiegazione teorica del fenomeno comunicativo.

Ma c'è di più. Grice costruisce la sua teoria delle massime conversazionali — quantità, qualità, relazione, modo — come principi che governano la cooperazione comunicativa tra parlanti razionali. Il Principio di Cooperazione afferma che i partecipanti a uno scambio verbale contribuiscono nella forma richiesta dallo scopo e dalla direzione dello scambio. Anche qui il presupposto è quello di un soggetto sufficientemente unificato da poter scegliere se e come rispettare le massime, e abbastanza trasparente a se stesso da sapere quando le sta violando. Quando un'organizzazione viola la massima di qualità producendo un comunicato deliberatamente fuorviante, la struttura della violazione è radicalmente diversa: non c'è necessariamente nessun singolo agente che intende ingannare; l'effetto ingannevole può emergere dal concatenarsi di contributi parziali ciascuno dei quali è soggettivamente onesto, o dall'assenza di un soggetto che abbia la visione d'insieme necessaria per riconoscere il carattere fuorviante del risultato.

Il *Grice implicaturale* — quello della logica conversazionale — produce l'immagine di un dialogo tra soggetti perfettamente razionali che inferiscono intenzioni comunicative sulla base di presupposizioni condivise circa la cooperatività dell'interlocutore. Ma questa immagine è una finzione normalizzante: il dialogo reale, sia nella comunicazione interpersonale che in quella istituzionale, è popolato di agenti la cui razionalità è limitata, le cui intenzioni sono parzialmente opache anche a se stessi, e la cui identità comunicativa è il prodotto di processi interni di negoziazione e conflitto che la teoria griceana non ha gli strumenti per modellare.

IV. Searle: istituzioni, intenzionalità collettiva e il nodo irrisolto

John Searle sviluppa la teoria degli atti linguistici di Austin in direzione sistematica, e ha il merito di affrontare esplicitamente la questione dell'intenzionalità collettiva e del linguaggio istituzionale. Il suo tentativo è il più ambizioso della tradizione classica, e le sue aporie sono perciò le più rivelatrici.

In *Speech Acts* Searle affina la tassonomia degli atti illocutori di Austin, introduce la distinzione tra regole costitutive e regole regolative, e consolida l'analisi delle condizioni di soddisfazione degli atti linguistici in termini di stati intenzionali del parlante. Un'asserzione ha come

contenuto intenzionale che una certa proposizione sia vera; una promessa ha come contenuto intenzionale che il parlante compirà una certa azione. Gli atti linguistici sono, in questa prospettiva, la forma linguistica dell'intenzionalità mentale: rappresentano nel medium pubblico del linguaggio gli stati interni di un soggetto psicologico.

Ma Searle ha anche l'acume di rendersi conto che questa psicologizzazione dell'atto linguistico porta a un problema serio con le istituzioni. In *The Construction of Social Reality* e poi in *Making the Social World*, egli sviluppa una teoria degli *status functions* e della *collective intentionality* per dare conto del fatto che le istituzioni — il denaro, il matrimonio, il governo, la proprietà privata — esistono in virtù di un accordo collettivo che non è riducibile alle intenzioni individuali dei singoli partecipanti. Il denaro è denaro perché noi — collettivamente, sistemicamente — lo trattiamo come tale; e questa realtà istituzionale è sostenuta da atti linguistici collettivi: le dichiarazioni che creano, modificano e distruggono realtà sociali.

La soluzione che Searle propone per il problema dell'intenzionalità collettiva è però, a ben guardare, elusiva. Egli afferma che l'intenzionalità collettiva è un fatto biologico primitivo — che gli esseri umani hanno la capacità di condividere intenzioni nel modo in cui "noi intendiamo fare X" non è riducibile a "io intendo fare X e credo che tu intenda fare X e credo che tu creda che io intenda fare X...". L'intenzione condivisa è, per Searle, un fenomeno psicologico reale che si realizza *in* ciascun individuo ma ha un contenuto irriducibilmente collettivo.

Questo è un punto sottile e importante, ma non risolve il problema che ci interessa. Anche ammettendo che ciascun individuo possa ospitare psicologicamente intenzioni a contenuto collettivo, resta la questione di come si produca e si sostenga nel tempo la convergenza tra le intenzioni collettive dei diversi individui che compongono un'istituzione. Searle tratta questa convergenza come un dato — le istituzioni funzionano quando i partecipanti condividono la stessa *deontic power*, lo stesso sistema di diritti e obblighi stabiliti dalle dichiarazioni istituzionali. Ma non offre un modello del processo attraverso cui questa convergenza si raggiunge, si mantiene o si perde. Non c'è spazio, nella sua teoria, per la conflittualità interna, per le intenzioni divergenti, per i processi di negoziazione che precedono e accompagnano ogni atto linguistico istituzionale reale.

La critica più acuta che si può rivolgere a Searle su questo punto riguarda la sua concezione delle *dichiarazioni* (declarations): quegli atti illocutori peculiari che non si limitano a rappresentare una realtà preesistente ma la creano attraverso la loro enunciazione. Le dichiarazioni sono, per Searle, l'asse portante della realtà istituzionale: è attraverso dichiarazioni che si fanno le leggi, si stipulano i contratti, si eleggono i rappresentanti, si fondano le istituzioni. Ma le dichiarazioni istituzionalmente rilevanti sono quasi sempre il prodotto di processi deliberativi collettivi — votazioni, trattative, procedure standardizzate — che coinvolgono agenti con intenzioni diverse e spesso conflittuali. La dichiarazione finale — "la legge è approvata", "il contratto è concluso", "l'imputato è condannato" — è il punto di convergenza di un processo distribuito, non l'espressione di un'intenzione unitaria.

V. Il soggetto come finzione regolatrice

La diagnosi che emerge da questa ricognizione è la seguente: Austin, Grice e Searle costruiscono le loro teorie assumendo implicitamente un modello di soggetto che è, nei casi tipici, una finzione regolatrice — utile, forse necessaria per certi scopi pratici, ma teoricamente distorsiva se elevata a fondamento dell'analisi.

La finzione del soggetto unitario funziona bene in un'area relativamente ristretta dell'uso linguistico: la conversazione faccia a faccia tra individui adulti, psicologicamente integrati, che interagiscono in condizioni di sostanziale parità e trasparenza. Anche qui, come vedremo, le neuroscienze cognitive e la psicologia dinamica hanno buone ragioni per mettere in discussione l'unitarietà del soggetto. Ma l'area di funzionamento della finzione si restringe drasticamente quando ci si sposta verso la comunicazione istituzionale, il discorso politico, il testo giuridico, i protocolli scientifici, le dichiarazioni aziendali — tutti quei contesti in cui l'atto linguistico è il prodotto di un processo distributivo che coinvolge molteplici agenti, regole procedurali, interessi divergenti e livelli di responsabilità non sovrapponibili.

La proposta teorica che svilupperemo nelle parti seguenti muove dalla constatazione che questi contesti non sono marginali rispetto all'uso linguistico totale: sono, per volume, importanza e conseguenze pratiche, il cuore del linguaggio sociale. I discorsi che governano la vita collettiva — le leggi, i trattati, i contratti, le politiche pubbliche, i comunicati

istituzionali, le sentenze, le risoluzioni delle organizzazioni internazionali — sono quasi tutti il prodotto di agenti distribuiti. Una teoria pragmatica adeguata non può trattarli come casi anomali o derivati rispetto al dialogo interpersonale; deve invece assumere il distribuito come paradigma principale e riconsiderare il caso del singolo parlante come un caso limite — importante, ma non fondante.

Questa inversione di prospettiva non è semplicemente un ampliamento del campo d'applicazione della teoria tradizionale. È una riconfigurazione teorica profonda, che richiede nuovi concetti e nuovi strumenti. Il concetto centrale di questa riconfigurazione — la *funzione di convergenza* — sarà introdotto e articolato nella terza parte. Prima, è necessario esaminare con maggiore precisione la natura dei sistemi distribuiti di cui ci occupiamo: le organizzazioni e le istituzioni, da un lato, e i sistemi cognitivi biologici, dall'altro. Perché la tesi più radicale di questa monografia è che la distribuzione non è una caratteristica eccezionale di certi enti collettivi ma la struttura normale di ogni sistema cognitivo capace di produrre atti linguistici — incluso il cervello umano individuale.

VI. Verso una genealogia del PSU

Prima di passare all'analisi dei sistemi distribuiti, vale la pena domandarsi da dove venga il PSU — quale contesto storico-culturale lo abbia reso così ovvio da sottrarlo all'esame critico. Una genealogia, anche sommaria, illumina perché rimpiazzarlo sia un'operazione filosoficamente non banale.

Il PSU affonda le radici in almeno tre tradizioni convergenti. La prima è quella della filosofia dell'azione di matrice aristotelica, che articola l'azione umana in termini di desiderio, credenza e deliberazione, e assegna al soggetto agente la capacità di integrare questi elementi in un'unità pratica — la *prohairesis*, la scelta deliberata. Questa tradizione ha attraversato interi secoli di filosofia morale e si è sedimentata nel senso comune colto come schema esplicativo dell'azione intenzionale. Un'azione è tale perché nasce dalla deliberazione di un agente; la deliberazione è il luogo in cui credenze e desideri si sintetizzano; la sintesi presuppone un soggetto unificato capace di essa.

La seconda tradizione è quella del contrattualismo moderno, da Hobbes a Locke a Rousseau, che costruisce la legittimità politica su atti comunicativi — il contratto sociale, il consenso, la volontà generale — il cui significato e validità dipendono dalla capacità dei firmatari di esprimere intenzioni autentiche e vincolanti. Il soggetto giuridico moderno — l'individuo capace di contrarre, promettere, obbligarsi — è pensato come unità psicologica sufficientemente stabile e trasparente a se stessa da poter essere ritenuta responsabile delle proprie dichiarazioni nel tempo.

La terza tradizione è quella dell'epistemologia cartesiana e della sua eredità, che situa la certezza nel cogito — nel punto di piena trasparenza della mente a se stessa — e costruisce l'impresa conoscitiva come un progetto individuale di ricognizione della propria interiorità e del mondo esterno. Questa epistemologia produce un soggetto che è, per definizione, il luogo della propria unità: la coscienza non può essere divisa perché la divisione implicherebbe l'impossibilità stessa del conoscere.

Queste tre tradizioni convergono nel produrre l'immagine di un soggetto che è insieme *agente* (delibera e sceglie), *giuridico* (si obbliga e risponde) e *epistémico* (conosce e si conosce). Questa immagine è così potente e così pervasiva da essere diventata il presupposto silenzioso non solo della pragmatica filosofica ma di larga parte delle scienze umane e sociali. Smontarla non significa negare che esista qualcosa che si può chiamare soggettività, o intenzionalità, o coscienza: significa rimettere in discussione l'idea che queste strutture abbiano la forma di un centro unificato, coerente e trasparente a se stesso.

Le neuroscienze cognitive, la psicologia della coscienza, le teorie dell'organizzazione e la filosofia dei sistemi complessi offrono, ciascuna a modo suo, ragioni per pensare che questa unità sia il prodotto — storicamente contingente, funzionalmente instabile, sempre parziale — di processi di coordinazione tra sotto-sistemi in competizione. Non il fondamento da cui partire, ma il risultato da spiegare. Ed è precisamente questo rovesciamento che apre la strada a una pragmatica dell'intenzione distribuita.

L'architettura della distribuzione

I. Cosa significa essere distribuiti

Prima di affrontare le conseguenze teoriche della distribuzione per la pragmatica, è necessario chiarire con precisione filosofica cosa si intende per sistema distribuito e perché la distribuzione è rilevante per la teoria degli atti linguistici. Il termine "distribuito" ha una storia tecnica nelle scienze cognitive e nell'informatica che vale la pena richiamare, ma occorre andare oltre la sua origine disciplinare per estrarne il potenziale filosofico.

Un sistema è distribuito, in senso minimale, quando le sue operazioni rilevanti non sono eseguibili da un singolo componente localizzato ma richiedono la coordinazione di molteplici componenti parzialmente autonomi. Questa definizione minima va però arricchita con una specificazione cruciale per i nostri scopi: un sistema è distribuito *intenzionalmente* quando non esiste in esso un unico componente che abbia accesso alla totalità delle informazioni rilevanti per determinare il contenuto e la forza degli atti linguistici che il sistema produce. La distribuzione epistemica — la frammentazione dell'informazione rilevante attraverso componenti diversi — è la caratteristica che più direttamente intacca i fondamenti della pragmatica classica.

Si può distinguere almeno tre forme di distribuzione che interessano la produzione degli atti linguistici. La prima è la *distribuzione agentiva*: più agenti concorrono alla produzione di un atto linguistico che nessuno di essi produce interamente da solo. È il caso del comunicato istituzionale redatto da un ufficio stampa sulla base di istruzioni della direzione, approvato da un comitato e firmato da un rappresentante legale. La seconda forma è la *distribuzione deliberativa*: la formazione dell'intenzione che guida l'atto linguistico è essa stessa il prodotto di un processo di negoziazione tra agenti con preferenze divergenti, e il risultato non è riducibile alle preferenze di nessun partecipante individuale. La terza è la *distribuzione cognitiva*: le risorse cognitive necessarie per produrre un atto linguistico — credenze, valutazioni, conoscenze contestuali, competenze procedurali — sono distribuite attraverso componenti diversi del sistema e non sono mai simultaneamente disponibili a un unico centro deliberativo.

Queste tre forme di distribuzione si trovano spesso congiunte nei casi reali, ma conviene distinguerle perché producono problemi teorici parzialmente diversi. La distribuzione agentiva pone il problema dell'identità del parlante: chi è il soggetto dell'atto linguistico? La

distribuzione deliberativa pone il problema dell'intenzione: quale intenzione ha prodotto l'atto? La distribuzione cognitiva pone il problema della competenza: l'atto è adeguatamente fondato sulle credenze e conoscenze necessarie per la sua riuscita?

II. Il cervello distribuito: la distribuzione come struttura normale della cognizione

La tesi che la distribuzione sia una caratteristica eccezionale di certi enti collettivi, da opporre alla unitarietà strutturale del singolo agente individuale, è indebolita — irrimediabilmente, si dirà — dalle scoperte delle neuroscienze cognitive degli ultimi tre decenni. Il cervello umano non è un sistema centralizzato con un esecutivo unico che integra le informazioni dei sotto-sistemi periferici e formula intenzioni coerenti. È, piuttosto, un sistema di sistemi parzialmente specializzati, parzialmente sovrapposti, che operano in parallelo, si inibiscono e si facilitano reciprocamente, e producono output comportamentali — inclusi gli atti linguistici — attraverso processi di competizione, selezione e coordinazione che non passano attraverso alcun centro deliberativo.

Le ricerche sul *split brain* inaugurate da Sperry e Gazzaniga negli anni Sessanta hanno per prime reso drammaticamente visibile questa struttura distribuita. Pazienti con il corpo calloso reciso per ragioni terapeutiche mostrano che i due emisferi possono ospitare credenze, motivazioni e giudizi di valore incompatibili, ciascuno capace di guidare comportamento verbale e non verbale in modo indipendente. Il cosiddetto *interprete dell'emisfero sinistro* — identificato da Gazzaniga — è un meccanismo che produce narrazioni post-hoc coerenti a partire da azioni generate da processi che l'emisfero stesso non controlla né conosce. La coerenza del soggetto che racconta se stesso è, in questa prospettiva, una costruzione ermeneutica che segue le azioni piuttosto che precederle come piano intenzionale.

Ma sarebbe sbagliato confinare queste considerazioni al caso patologico del cervello diviso. La ricerca sul *processing* inconscio — dai lavori pionieristici di Libet sull'attività cerebrale che precede la consapevolezza dell'intenzione motoria, agli studi più recenti sulla cognizione non conscia — suggerisce che larga parte dell'attività psicologica che sostiene la produzione linguistica si svolge al di fuori della coscienza e non è accessibile alla deliberazione esplicita. Quando parliamo, attingiamo a

risorse fonologiche, morfosintattiche, pragmatiche e semantiche che operano in modo automatico e parallelo, la cui coordinazione produce l'enunciato finale senza che il parlante abbia accesso diretto ai processi sottostanti.

Per la pragmatica questo ha conseguenze non banali. Se l'intenzione comunicativa — nel senso griceano di intenzione a essere riconosciuta come tale — è essa stessa il prodotto di processi inconsci in parte inaccessibili al parlante, allora l'immagine del parlante che *ha* un'intenzione e *sceglie* di esprimerla linguisticamente è una semplificazione che cattura solo la superficie fenomenologica di un processo molto più complesso. Il parlante non è il luogo da cui origina l'intenzione: è, piuttosto, il sistema attraverso cui una varietà di processi sub-personali converge in un output linguistico che può essere retroattivamente descritto in termini intenzionali.

La psicologia dinamica — da Freud in poi, con tutte le sue ramificazioni e trasformazioni — ha prodotto una riflessione sulla struttura non unificata del soggetto che va nella stessa direzione. Il conflitto intrapsichico, l'ambivalenza, la scissione, la dissociazione: questi fenomeni indicano che il soggetto psicologico non è un sistema con un unico centro intenzionale ma una confederazione di istanze parzialmente autonome che possono volere cose diverse, parlare con voci diverse, e produrre atti linguistici la cui "vera" intenzione è opaca allo stesso parlante. La psicoanalisi ha elaborato una sofisticata teoria degli atti linguistici — i lapsus, i sogni, i sintomi come "formazioni di compromesso" tra istanze in conflitto — che è implicitamente una pragmatica del distribuito, anche se non l'ha mai articolata in questi termini.

Non si tratta, qui, di abbracciare acriticamente né il riduzionismo neurologico né il modello psicoanalitico. Si tratta di riconoscere che entrambe le tradizioni convergono nell'indicare che l'unitarietà del soggetto psicologico è un'approssimazione — funzionalmente adeguata per molti scopi ordinari, ma teoricamente inadeguata come fondamento di una teoria pragmatica generale. Il soggetto individuale è già, in senso preciso, un sistema distribuito: la sua coerenza apparente è il prodotto di processi di integrazione che possono fallire, vacillare o produrre conflitti manifesti.

III. Le organizzazioni come sistemi linguistici distribuiti

Se il singolo soggetto è già un sistema distribuito, le organizzazioni sono sistemi distribuiti di sistemi distribuiti, con conseguente amplificazione di tutte le difficoltà teoriche che abbiamo identificato. Un'organizzazione — un'azienda, un'università, un partito politico, un'agenzia governativa, una chiesa — produce una quantità enorme di atti linguistici: contratti, delibere, comunicati, dichiarazioni di politica, normative interne, corrispondenza ufficiale, posizioni pubbliche. Ciascuno di questi atti ha potenzialmente forza giuridica, morale e pragmatica. Ma la loro produzione è il risultato di processi organizzativi che coinvolgono molteplici agenti con ruoli, informazioni e interessi diversi.

La teoria delle organizzazioni ha sviluppato strumenti concettuali per descrivere questi processi: la teoria della razionalità limitata di Simon, l'analisi delle coalizioni di Cyert e March, la teoria dell'agenzia, i modelli di garbage can decision-making. Questi strumenti mostrano che le decisioni organizzative — incluse quelle che danno origine ad atti linguistici con effetti esterni — raramente emergono da un processo deliberativo unitario che ottimizza su preferenze coerenti. Emergono invece da processi che combinano vincoli procedurali, asimmetrie informative, pressioni temporali, conflitti di interesse e dinamiche di potere in modi che possono produrre output incoerenti, contraddittori o non anticipati da nessuno dei partecipanti.

Si consideri il processo che porta alla pubblicazione di una *policy* aziendale sulla privacy dei dati. Il testo finale è prodotto dalla combinazione di contributi del reparto legale (che massimizza la conformità normativa), del marketing (che massimizza la possibilità di raccogliere dati per profilazione), dell'IT (che traduce vincoli tecnici in linguaggio di policy), della direzione (che media tra obiettivi commerciali e rischio reputazionale), e degli uffici di comunicazione (che adattano il linguaggio per la comprensibilità pubblica). Ciascun contribuente ha una propria intenzione comunicativa parziale; il testo finale è una soluzione di compromesso che non rispecchia pienamente l'intenzione di nessuno. Eppure questo testo ha forza di dichiarazione vincolante: chi lo legge ha diritto di aspettarsi che l'organizzazione si comporti in conformità con quanto dichiarato.

La questione pragmatica fondamentale è: qual è l'intenzione comunicativa di questa dichiarazione? Nel senso griceano, l'intenzione comunicativa è quella di produrre nell'interlocutore determinate credenze

e aspettative circa il comportamento dell'organizzazione. Ma quale sotto-sistema organizzativo è il portatore di questa intenzione? Il reparto legale che ha redatto la versione preliminare? La direzione che l'ha approvata? Il comunicatore che l'ha pubblicata? Nessuna di queste risposte è soddisfacente, perché nessuno di questi agenti ha la visione completa dei processi che hanno portato al testo finale, né è pienamente responsabile del suo contenuto.

Questa frammentazione dell'intenzione non è un'imperfezione contingente del processo organizzativo, rimediabile con migliore coordinazione o comunicazione interna. È strutturale: qualunque sistema di dimensioni non banali che produce atti linguistici con forza esterna deve necessariamente distribuire le competenze e le responsabilità relative alla loro produzione. La specializzazione — che è la fonte principale dell'efficienza organizzativa — porta inevitabilmente con sé la distribuzione dell'intenzione comunicativa.

IV. Le istituzioni come macchine performative distribuite

Le istituzioni sono un caso ancora più interessante delle organizzazioni, perché la loro funzione primaria è precisamente quella di produrre atti linguistici con forza costitutiva: dichiarazioni che creano, modificano e distruggono realtà giuridiche e sociali. Un parlamento approva leggi; un tribunale emette sentenze; un ufficio del registro riceve e certifica contratti; una banca centrale emette moneta e regola tassi d'interesse. Tutti questi sono atti linguistici nel senso tecnico della tradizione pragmatica — hanno forza illocutoria, condizioni di soddisfazione, effetti perlocutori — ma nessuno di essi ha un singolo titolare intenzionale.

La dimensione performativa delle istituzioni è ciò che le rende filosoficamente più urgenti del semplice caso organizzativo. Un'azienda che produce un comunicato commerciale produce certamente un atto linguistico, ma le sue conseguenze pratiche dipendono in ultima istanza da come i destinatari lo recepiscono e se gli danno credito. Una sentenza penale, per contro, ha forza performativa indipendente dalle credenze dei destinatari: l'imputato è condannato anche se non lo accetta, la proprietà passa all'erede anche se lo contesta. La forza istituzionale degli atti linguistici non dipende dal riconoscimento dei destinatari ma dall'autorità costitutiva del sistema istituzionale che li produce.

Searle ha analizzato con cura questa dimensione, ma — come si è già notato — la sua analisi presuppone un sistema istituzionale in cui le *deontic powers* siano chiaramente assegnate e le dichiarazioni istituzionalmente valide emergano senza conflitti dalla corretta applicazione delle regole costitutive. Nella realtà, le istituzioni funzionano in modo molto più caotico e conflittuale. La produzione di una legge è il risultato di un processo legislativo che coinvolge centinaia di agenti con interessi e visioni diverse; il testo finale è spesso ambiguo, contraddittorio o lacunoso proprio perché è stato il campo di battaglia di forze in conflitto che hanno lasciato tracce nella forma finale del testo. La sentenza di un tribunale collegiale è prodotta dalla deliberazione di giudici con interpretazioni diverse del diritto applicabile; il dispositivo finale può essere il risultato di negoziazioni tra posizioni divergenti, e il soggetto dell'atto linguistico "la corte condanna" è un sistema distribuito i cui componenti possono dissentire — come dimostrano le opinioni dissenzienti — pur vincolandosi collettivamente all'output.

Ciò che rende le istituzioni peculiarmente interessanti dal punto di vista della pragmatica distribuita è che esse producono atti linguistici la cui forza non dipende dall'intenzione dei loro produttori ma dalla posizione che occupano nella struttura istituzionale. Una sentenza pronunciata da un giudice competente ha forza di cosa giudicata indipendentemente dalle intenzioni psicologiche del giudice: anche un giudice che abbia sbagliato la propria valutazione o che non abbia pienamente compreso le implicazioni della propria decisione produce un atto linguistico vincolante. Questo è esattamente il tipo di fenomeno che la teoria classica non è attrezzata per descrivere, perché separa nettamente la validità dell'atto dalla sua forza intenzionale.

V. Le comunità linguistiche come sistemi distribuiti di normatività

C'è ancora un livello, forse il più fondamentale, in cui la distribuzione è strutturalmente necessaria: quello della comunità linguistica come tale. Il linguaggio non è una risorsa che i parlanti portano con sé come una borsa degli attrezzi; è un sistema di norme condivise che preesistono a ogni parlante individuale, lo attraversano e lo sopravvivono. Questa ovvietà ha conseguenze per la pragmatica che non sono state sufficientemente esplorate.

Le convenzioni linguistiche — semantiche, sintattiche, pragmatiche — che rendono possibili gli atti linguistici sono il prodotto di una storia comunitaria che nessun parlante ha controllato e che nessun parlante può modificare unilateralmente. Quando uso una parola, attivo un campo di significati e connotazioni che è il sedimento di milioni di usi precedenti da parte di milioni di parlanti. Quando compio un atto illocutorio, mi inserisco in una struttura di pratiche linguistiche che non ho creato e che vincola il mio atto indipendentemente dalle mie intenzioni. In questo senso, ogni atto linguistico individuale è già il prodotto di un agente distribuito: il parlante individuale è solo il punto di passaggio attuale di una corrente linguistica che ha la sua inerzia propria.

Wittgenstein aveva intuito questo con la sua critica al linguaggio privato e con la nozione di *Lebensform*: il linguaggio è possibile solo dentro una forma di vita condivisa, e le regole linguistiche non sono decisioni individuali ma pratiche collettive che si riproducono attraverso l'uso. Ma Wittgenstein non ha sviluppato una pragmatica sistematica a partire da questa intuizione. Ha invece smontato la possibilità di fondare il significato sull'intenzione privata, senza costruire in positivo una teoria di come il significato funzioni in sistemi distribuiti.

La proposta che svilupperemo nella parte successiva cerca di raccogliere questo lascito senza rinunciare alla dimensione normativa e pragmatica che Wittgenstein, nella sua fedeltà all'uso ordinario, non ha mai formalizzato.

VI. La distribuzione nei sistemi artificiali: verso una pragmatica non biologica

Un ultimo tipo di agente distribuito merita attenzione, perché rappresenta sia il caso estremo della nostra riflessione sia la sfida più urgente per le applicazioni pratiche della pragmatica contemporanea: i sistemi di intelligenza artificiale basati su architetture distribuite.

I modelli linguistici di grandi dimensioni — i cosiddetti *large language models* — producono atti linguistici in senso funzionalmente preciso: generano testi che hanno forza assertiva, che fanno promesse implicite (di accuratezza, di completezza, di pertinenza), che danno consigli, che emettono giudizi, che assumono posizioni. Eppure questi sistemi non hanno, per definizione, un centro intenzionale: sono reti di miliardi di

parametri il cui comportamento emerge dalla coordinazione di processi distribuiti senza alcuna architettura centralizzata che ospiti qualcosa che assomigli a un'intenzione cosciente.

Applicare la pragmatica classica a questi sistemi è impossibile in modo non troppo diverso da come è impossibile applicarla a un'istituzione: non c'è un parlante con uno stato mentale che causa l'atto linguistico. Eppure sarebbe assurdo negare che questi sistemi producano atti linguistici: i loro output hanno forza illocutoria, condizioni di soddisfazione, effetti perlocutori, e — come dimostrano le controversie giuridiche in corso in molte giurisdizioni — conseguenze pratiche e morali reali.

La pragmatica dell'intenzione distribuita è necessaria, tra l'altro, per dare una risposta adeguata alla questione di come debbano essere descritti, valutati e — soprattutto — a chi debbano essere ascritti gli atti linguistici prodotti da sistemi artificiali. Una risposta adeguata non può essere fornita né dalla teoria classica (che richiede un soggetto intenzionale) né da una posizione eliminativista (che nega che questi sistemi producano atti linguistici in senso proprio), ma richiede un apparato concettuale nuovo — precisamente quello che questa monografia si propone di abbozzare.

La funzione di convergenza — verso una teoria degli atti linguistici per agenti distribuiti

---

I. Il problema della sostituzione

L'impresa teorica che si tenta in questa parte è delicata e per certi versi paradossale: si tratta di costruire una teoria degli atti linguistici che mantenga la ricchezza descrittiva e la forza normativa della tradizione classica, rinunciando però al suo fondamento intenzionalista. Il pericolo principale è quello di perdere troppo: una teoria che si limiti a descrivere il comportamento linguistico senza più poter dire nulla di cogente su quando un atto linguistico *riesce*, quando *obbliga*, quando è *responsabile* di certe conseguenze, sarebbe pragmaticamente inutile nel senso più letterale.

Il punto di partenza è il riconoscimento che l'intenzione svolge nella teoria classica non uno ma almeno tre ruoli distinti, e che questi ruoli

non hanno necessariamente bisogno di essere svolti dallo stesso concetto sostitutivo. L'intenzione nella pragmatica classica serve, primo, come *principio di individuazione* dell'atto linguistico: è la struttura intenzionale del parlante che determina quale tipo di atto illocutorio viene compiuto (una promessa, non un'asserzione; un comando, non una preghiera). Secondo, l'intenzione serve come *condizione di felicità*: l'atto riesce solo se l'intenzione che lo anima è appropriata al tipo di atto e alle circostanze. Terzo, l'intenzione serve come *principio di responsabilità*: è perché l'atto nasce da un'intenzione attribuibile al parlante che il parlante può essere considerato responsabile delle sue conseguenze.

Una pragmatica per agenti distribuiti deve trovare sostituti adeguati per ciascuno di questi tre ruoli. La proposta che avanzo è che tutti e tre possano essere ricoperti da un unico concetto, più ricco e più complesso dell'intenzione individuale: la *funzione di convergenza*.

II. La funzione di convergenza: definizione formale e intuitiva

In modo intuitivo, la funzione di convergenza è il processo attraverso cui un sistema distribuito di agenti con preferenze, informazioni e posizioni parzialmente divergenti produce un output comunicativo che funziona pragmaticamente come se fosse il prodotto di un'intenzione unitaria. Non si tratta di ridurre la molteplicità dei contribuenti a un'unità fittizia: si tratta di identificare il meccanismo che trasforma la molteplicità in un output dotato di forza illocutoria determinata, condizioni di soddisfazione specifiche ed effetti perlocutori prevedibili.

Più formalmente, si può definire la funzione di convergenza come segue. Sia $\mathcal{A} = {a_1, a_2, ..., a_n}$ un insieme di agenti che concorrono alla produzione di un atto linguistico. Ciascun agente a_i porta al processo un vettore v_i che comprende: le sue preferenze circa il contenuto proposizionale dell'atto; le sue preferenze circa la forza illocutoria; le sue credenze rilevanti circa il contesto; le sue aspettative circa gli effetti perlocutori; i suoi vincoli procedurali (il ruolo che occupa nel processo produttivo). La funzione di convergenza F è una mappatura da ${v_1, ..., v_n}$ a un output comunicativo O che ha forza illocutoria determinata, contenuto proposizionale determinato, e condizioni di soddisfazione determinate.

Questa definizione non presuppone che F sia una funzione semplice, né che il suo risultato soddisfi equamente tutte le preferenze dei partecipanti. F può essere asimmetrica (alcune voci pesano più di altre), può essere dipendente dal sentiero (l'ordine in cui i contributi vengono integrati conta), può essere sensibile a vincoli procedurali che non coincidono con le preferenze di nessun agente, può produrre output che nessun agente avrebbe individualmente scelto. Ciò che è richiesto è solo che il risultato della convergenza abbia le proprietà pragmatiche necessarie perché possa funzionare come atto linguistico in senso pieno.

Il vantaggio concettuale di questo approccio è che consente di *descrivere la struttura della convergenza* — di chiedersi come è fatta F in un determinato caso — senza presupporre che il risultato sia la volontà unificata di un soggetto. La struttura di F può essere gerarchica (il capo decide dopo aver consultato i collaboratori), procedurale (la maggioranza vince), negoziale (il testo finale è il risultato di concessioni reciproche), o algoritmica (le regole del sistema determinano l'output senza che nessun agente lo voglia esplicitamente). Ogni struttura di convergenza diversa produce condizioni di felicità diverse, determina diversamente i casi di fallimento dell'atto, e distribuisce diversamente la responsabilità tra i contribuenti.

III. Forza illocutoria senza intenzione: il ruolo della struttura di convergenza

Uno degli interrogativi più delicati per la pragmatica del distribuito è quello di come si determini la forza illocutoria di un atto linguistico prodotto da un sistema distribuito. Nella teoria classica, è la struttura intenzionale del parlante che determina la forza: produco una promessa se ho l'intenzione di obbligarmi a un'azione futura, produco un'asserzione se ho l'intenzione di indurre nell'interlocutore la credenza che una certa proposizione è vera, e così via. Ma nella pragmatica distribuita, non c'è un centro intenzionale che possa essere interrogato per determinare la forza.

La risposta che propongo è che la forza illocutoria di un atto linguistico prodotto da un sistema distribuito è determinata dalla *struttura di convergenza* che lo ha generato e dal *contesto istituzionale* in cui il sistema opera. Questi due fattori sostituiscono congiuntamente il ruolo dell'intenzione individuale nella determinazione della forza.

La struttura di convergenza è rilevante perché determina quale tipo di impegno il sistema prende collettivamente. Un delibera approvata all'unanimità da un consiglio ha una forza illocutoria diversa da una dichiarazione emessa in via d'urgenza da un singolo amministratore delegato con delega limitata: nel primo caso il sistema ha mobilizzato le sue procedure massimali di convergenza, producendo un impegno più robusto; nel secondo il sistema ha delegato la convergenza a un sotto-sistema con competenza limitata, producendo un impegno più circoscritto. Queste differenze hanno conseguenze dirette sulle condizioni di soddisfazione dell'atto e sulla sua forza normativa.

Il contesto istituzionale è rilevante perché assegna forze illocutorie specifiche agli output di certi sistemi distribuiti indipendentemente da qualunque intenzione dei partecipanti. Una sentenza pronunciata da un tribunale competente ha forza illocutoria di cosa giudicata indipendentemente dall'intenzione dei giudici che la compongono. Una legge approvata seguendo le procedure costituzionali ha forza normativa anche se nessun parlamentare aveva compreso pienamente le sue implicazioni. In questi casi, la forza illocutoria è assegnata dalla struttura istituzionale al prodotto del processo di convergenza, non ai componenti individuali del processo.

Questo meccanismo — che potremmo chiamare *forza illocutoria procedurale* — è la caratteristica più distintiva degli atti linguistici istituzionali e quella che li differenzia più nettamente dagli atti linguistici individuali. La forza procedurale è robusta rispetto alle intenzioni dei singoli contribuenti: anche se tutti i parlamentari avessero votato per ragioni diverse, anche se nessuno avesse pienamente compreso il testo della legge, la legge ha la forza che le assegna la procedura costituzionale. Questo non vuol dire che le intenzioni individuali siano irrilevanti — possono essere rilevanti per l'interpretazione del testo, per la valutazione della sua legittimità sostanziale, per la determinazione degli effetti perlocutori. Ma non sono costitutive della forza illocutoria.

IV. Condizioni di felicità per sistemi distribuiti

La teoria classica delle condizioni di felicità degli atti linguistici specifica quando un atto riesce o fallisce in riferimento alle intenzioni, credenze e capacità del parlante. Searle, nella sua sistemazione, distingue tra condizioni preparatorie, condizioni di sincerità e condizioni essenziali.

Per una promessa, ad esempio: condizione preparatoria che il parlante possa compiere l'azione promessa e che l'interlocutore preferisca che la compia; condizione di sincerità che il parlante intenda davvero compierla; condizione essenziale che il parlante si obblighi a compierla.

Per una pragmatica del distribuito, queste condizioni devono essere riformulate in modo da non presupporre un soggetto unitario. La proposta è di distinguere tra condizioni di felicità *strutturali* e condizioni di felicità *sostanziali*.

Le condizioni strutturali riguardano il processo di convergenza: l'atto riesce strutturalmente quando il processo che lo ha prodotto è conforme alle procedure che il sistema ha adottato per produrre atti di quel tipo. Una delibera del consiglio di amministrazione riesce strutturalmente se è stata adottata con il quorum necessario, nella sede prevista, con le modalità di convocazione corrette. Una sentenza riesce strutturalmente se è stata pronunciata da un giudice competente, nel rispetto delle garanzie processuali, nelle forme prescritte. Le condizioni strutturali sono, in sostanza, le condizioni di *validità procedurale* dell'atto.

Le condizioni sostanziali riguardano invece il contenuto del processo di convergenza: l'atto riesce sostanzialmente quando i contribuenti al processo di convergenza avevano le credenze, le competenze e le informazioni necessarie per produrre un atto del tipo in questione. Una promessa istituzionale riesce sostanzialmente se il sistema che la fa è in grado di mantenerla e ha considerato seriamente le condizioni della sua esecuzione. Un'asserzione istituzionale riesce sostanzialmente se il sistema che la produce dispone delle informazioni necessarie per sapere che è vera. Le condizioni sostanziali corrispondono, nella pragmatica distribuita, alla condizione di sincerità della teoria classica — ma riformulate in termini di accesso informativo del sistema piuttosto che di stati mentali del parlante.

Il punto critico è che le condizioni strutturali e quelle sostanziali possono dissociarsi. Un atto può essere strutturalmente valido ma sostanzialmente fallace: una delibera può essere formalmente corretta ma basata su informazioni erronee che nessun membro del consiglio aveva interesse a correggere. Viceversa, un atto può essere sostanzialmente ben fondato ma strutturalmente invalido: un funzionario con conoscenza dettagliata di un problema può fare una promessa pubblica che non aveva la competenza formale di fare, ma che il sistema avrebbe certamente confermato se interpellato. Le conseguenze pratiche e giuridiche di questi

casi di dissociazione sono profonde e sarà necessario esaminarle con cura nella quarta parte.

V. La teoria dei sotto-sistemi in competizione

Un'ulteriore specificazione della funzione di convergenza riguarda la dinamica interna del processo che la realizza. In molti sistemi distribuiti, la produzione di un atto linguistico non è il risultato di una semplice aggregazione di preferenze ma di una competizione tra sotto-sistemi con interessi parzialmente divergenti. Questa competizione può risolversi in modo diverso a seconda della struttura di potere interna al sistema, e il suo esito determina quali intenzioni parziali prevalgono nella forma finale dell'atto.

Si consideri il caso del sistema giudiziario nella produzione di giurisprudenza. Ogni sentenza è il punto di convergenza di pressioni provenienti da sotto-sistemi diversi: la logica interna del precedente (che spinge verso la coerenza con le decisioni passate), le richieste delle parti (che spingono verso la risoluzione del caso specifico), le istanze di giustizia sostanziale percepite dal giudicante (che possono spingere verso l'equità anche a costo della coerenza formale), le pressioni sistemiche (l'adattamento del diritto ai cambiamenti sociali). Il testo della sentenza è la traccia del processo di convergenza tra questi sotto-sistemi, e la sua interpretazione richiede la ricostruzione di questo processo piuttosto che l'attribuzione di un'intenzione unitaria al "giudice".

La teoria dei sotto-sistemi in competizione consente di modellare fenomeni pragmatici che la teoria classica non può descrivere: l'ambiguità intenzionale (quando il testo è il prodotto di un compromesso tra posizioni incompatibili), la policromia illocutoria (quando un atto ha forze illocutorie diverse per i diversi sotto-sistemi che vi hanno contribuito), e l'instabilità della forza (quando il risultato del processo di convergenza è fragile e potrebbe cambiare al mutare delle condizioni di potere interne al sistema).

L'ambiguità intenzionale è particolarmente rilevante perché costituisce una categoria pragmatica genuinamente nuova rispetto alla tradizione classica. Nella pragmatica tradizionale, l'ambiguità è un fenomeno principalmente semantico (un enunciato ha più interpretazioni linguistiche) o contestuale (le circostanze non determinano

univocamente la forza illocutoria). L'ambiguità intenzionale — che l'atto sia il prodotto di intenzioni parziali incompatibili che non si sono pienamente integrate — è invece un fenomeno strutturale specifico degli atti linguistici distribuiti. Un testo normativo che contiene disposizioni contraddittorie non è semplicemente "ambiguo" in senso linguistico: è la traccia di un conflitto non risolto tra sotto-sistemi in competizione, e la sua interpretazione richiede la ricostruzione di questo conflitto piuttosto che la ricerca di un'intenzione unica del legislatore.

VI. La dimensione temporale della convergenza

Un aspetto che la definizione formale della funzione di convergenza non cattura pienamente è la sua dimensione temporale. La convergenza non è un evento puntuale ma un processo che si svolge nel tempo, e questo ha importanti conseguenze per la teoria degli atti linguistici distribuiti.

Prima di tutto, il processo di convergenza ha una storia: i contributi dei diversi agenti si stratificano nel tempo, e la struttura del processo può cambiare durante il suo svolgimento. Un trattato internazionale negoziato per anni riflette, nel suo testo finale, la storia delle negoziazioni: le concessioni fatte in momenti diversi, i compromessi raggiunti in certe fasi e poi parzialmente riaperti in altre, le asimmetrie di potere che si sono spostate nel corso del processo. Il testo finale non è la somma sincronicistica di preferenze attuali dei firmatari ma il sedimento di un processo diacronico.

In secondo luogo, la convergenza può essere solo parziale e provvisoria. Un sistema distribuito può produrre atti linguistici che rappresentano stati di equilibrio instabile tra sotto-sistemi in conflitto, e questi equilibri possono rompersi al cambiare delle condizioni. Una politica aziendale adottata sotto una certa direzione può essere tacitamente non rispettata dopo un cambio ai vertici; un accordo internazionale può essere eroso dall'interpretazione divergente che le parti ne danno nel tempo. La forza di un atto linguistico distribuito non è quindi una proprietà stabile che si acquista al momento della produzione: è una proprietà dinamica che dipende dal continuo processo di riproduzione della convergenza nel tempo.

Questa dimensione temporale introduce nella pragmatica del distribuito una categoria che la teoria classica non aveva ragione di considerare:

quella di *atto linguistico in sospeso*, un atto prodotto da una convergenza provvisoria la cui forza dipende dal mantenimento di un equilibrio che potrebbe rompersi. Le costituzioni degli stati, i trattati di pace, i grandi accordi commerciali sono tutti atti linguistici in questo senso: la loro forza normativa è reale e operante, ma dipende dal continuo consenso attivo dei sistemi distribuiti che li hanno prodotti e che ne incarnano l'esecuzione.

VII. Implicature e inferenza in sistemi distribuiti

L'ultimo aspetto teorico che occorre affrontare in questa parte riguarda il meccanismo delle implicature. La logica conversazionale griceana descrive come i destinatari di un messaggio inferiscano, a partire dal contenuto esplicito dell'enunciato e dall'assunzione di cooperatività del parlante, significati impliciti che non sono parte del contenuto semantico letterale. Questo meccanismo è fondamentale per la comprensione del linguaggio ordinario, e qualunque pragmatica del distribuito deve dare conto di come funzioni in contesti non interpersonali.

Il problema è che l'assunzione di cooperatività — il pilastro del meccanismo griceano — ha un senso chiaro quando si applica a un parlante individuale (posso assumere che stia cercando in buona fede di comunicarmi qualcosa), ma diventa molto più problematica quando si applica a un sistema distribuito i cui componenti hanno interessi parzialmente divergenti. Un comunicato aziendale non è prodotto da un agente cooperativo in senso griceano: è il prodotto di una convergenza tra sotto-sistemi di cui alcuni possono avere interessi comunicativi che non coincidono con quelli del destinatario.

La proposta è di sostituire l'assunzione di cooperatività con quella che potremmo chiamare l'*assunzione di stabilità funzionale*: il destinatario di un atto linguistico distribuito può razionalmente assumere che il sistema che lo ha prodotto abbia un interesse funzionale nel mantenere la coerenza tra i propri atti linguistici e le proprie pratiche, non per ragioni di benevolenza verso il destinatario, ma per ragioni di autoriproduzione istituzionale. Un'istituzione che sistematicamente viola i propri impegni perde credibilità e capacità di azione; un'organizzazione che produce comunicati deliberatamente ingannevoli accumula passività reputazionali. L'assunzione di stabilità funzionale è più debole e più cinica

dell'assunzione di cooperatività, ma è più realistica e applicabile ai sistemi distribuiti.

Le implicature conversazionali nel contesto distribuito sono quindi inferenze che il destinatario trae assumendo non la buona fede del parlante ma la razionalità funzionale del sistema. "L'organizzazione X ha dichiarato che il prodotto Y è sicuro" implica conversazionalmente, in condizioni normali, che l'organizzazione dispone di evidenze che supportano questa asserzione, non perché si assuma che l'organizzazione voglia il bene del destinatario, ma perché un'asserzione pubblica senza evidenze sarebbe incompatibile con gli interessi di autoriproduzione dell'organizzazione. Quando questi interessi cambiano — quando il costo di ritirare la dichiarazione supera quello di mantenerla — l'implicatura si incrina, e con essa la struttura comunicativa dell'atto.

Conseguenze — Responsabilità, istituzioni e politica del linguaggio

---

I. Il problema della responsabilità distribuita

Nessuna delle questioni pratiche che emergono dalla pragmatica dell'intenzione distribuita è più pressante di quella della responsabilità. Le teorie giuridiche e morali della responsabilità sono costruite, nella loro grande maggioranza, intorno al modello del soggetto unitario: si risponde di ciò che si è inteso fare, di ciò che si poteva prevedere, di ciò che si aveva la capacità di evitare. Quando l'atto linguistico è il prodotto di un sistema distribuito, ognuna di queste condizioni diventa problematica in modi che il diritto e la filosofia morale tradizionali non sono pienamente attrezzati a gestire.

Partiamo dalla responsabilità giuridica del discorso — la responsabilità per atti linguistici che producono danno. Nel diritto italiano, come in molti sistemi di common law e civil law, la responsabilità per dichiarazioni false o fuorvianti presuppone tipicamente l'elemento soggettivo: la conoscenza della falsità (nel caso della diffamazione), l'intenzione fraudolenta (nel caso della truffa contrattuale), la colpa grave per omissione di controlli (nel caso della responsabilità per informazioni erronee nei mercati finanziari). Tutti questi criteri riconducono la

responsabilità a un soggetto psicologico che ha o avrebbe dovuto avere una certa struttura mentale.

Quando il soggetto è un'organizzazione, il diritto adotta tipicamente soluzioni di compromesso che non risolvono il problema teorico. La responsabilità amministrativa degli enti (in Italia, il d.lgs. 231/2001) attribuisce responsabilità all'ente per reati commessi "nell'interesse o a vantaggio" dell'ente da persone che rivestono funzioni apicali o sono sottoposte alla loro direzione. Questa soluzione è pragmaticamente efficace come deterrente, ma teoricamente insoddisfacente perché trasferisce il problema: se l'atto linguistico illecito è il prodotto di una convergenza distribuita in cui nessun singolo agente aveva il pieno controllo del risultato, attribuire la responsabilità ai soli "apicali" è una finzione che punisce il rango piuttosto che la causazione.

La pragmatica dell'intenzione distribuita suggerisce una diversa architettura della responsabilità giuridica del discorso, articolata in tre livelli correlati alla struttura della funzione di convergenza.

Al primo livello vi è la *responsabilità strutturale*: l'ente risponde per gli atti linguistici prodotti da processi di convergenza che erano sotto il suo controllo normativo, indipendentemente dall'intenzione di singoli partecipanti. Questa responsabilità è più ampia di quella tradizionale perché non richiede l'identificazione di un responsabile individuale, ma è anche più precisa perché riguarda specificamente la struttura del processo piuttosto che il rango dei suoi partecipanti.

Al secondo livello vi è la *responsabilità epistemica distribuita*: i singoli partecipanti al processo di convergenza rispondono in proporzione alla loro posizione epistemica — a ciò che potevano e dovevano sapere circa le conseguenze dell'atto linguistico finale. Un consigliere che ha approvato una delibera senza leggerne le implicazioni non è responsabile allo stesso modo di chi le ha redatte sapendo che erano fuorvianti; ma non è nemmeno completamente esente dalla responsabilità, perché il suo contributo al processo di convergenza ha reso possibile l'atto illecito.

Al terzo livello vi è la *responsabilità per design della convergenza*: coloro che progettano e mantengono le strutture attraverso cui il processo di convergenza si realizza — i legislatori che scrivono le procedure, i dirigenti che definiscono i protocolli, i professionisti che disegnano i sistemi di governance — rispondono per gli effetti sistemici di queste strutture sugli atti linguistici che producono. Se una struttura di

governance è progettata in modo da rendere impossibile la verifica delle informazioni che alimentano gli atti linguistici dell'ente, chi ha disegnato questa struttura porta una responsabilità per i danni derivanti da atti linguistici epistemicamente mal fondati.

II. La verità nelle dichiarazioni istituzionali

Una delle questioni filosoficamente più profonde che la pragmatica del distribuito apre riguarda le condizioni di verità delle asserzioni istituzionali. Quando un governo afferma che l'economia cresce, quando un'agenzia di rating attribuisce a uno stato sovrano la tripla A, quando un'autorità sanitaria dichiara che un farmaco è sicuro ed efficace, si producono asserzioni con condizioni di verità apparentemente ordinarie: sono vere se i fatti corrispondono al contenuto proposizionale. Ma la struttura distribuita del soggetto asserente introduce complicazioni che non sono riducibili al semplice problema della verifica fattuale.

Il problema fondamentale è che la verità di un'asserzione presuppone, nella pragmatica tradizionale, che il parlante abbia una credenza corrispondente al contenuto proposizionale asserito. Questa condizione — la condizione di sincerità di Searle — diventa problematica per un'asserzione istituzionale perché non esiste un singolo luogo nel sistema dove si possa localizzare questa credenza. I diversi sotto-sistemi che hanno contribuito alla produzione dell'asserzione possono avere credenze diverse e persino incompatibili circa il suo contenuto.

Si consideri un caso reale: il governo di un paese produce un rapporto ufficiale che afferma che le emissioni di gas serra sono diminuite del 15% rispetto all'anno precedente. Il ministero dell'ambiente ha dati che supportano questa cifra; il ministero dell'economia ha internamente sollevato dubbi sulla metodologia di calcolo che sono stati silenziati; il dipartimento statistico ha prodotto una stima alternativa che indica una diminuzione del 7%. La dichiarazione ufficiale del governo è *sinceramente falsa*? È *menzognera*? È *fuorviante*? O è semplicemente il prodotto di un processo di convergenza che ha selezionato una certa metodologia di calcolo tra quelle disponibili?

La risposta che propone la pragmatica del distribuito è che queste categorie — verità, falsità, sincerità, menzogna — devono essere ridefinite in termini che si adattino alla struttura del sistema che produce

l'asserzione. Un'asserzione istituzionale è *epistemicamente giustificata* quando il processo di convergenza che l'ha prodotta ha mobilizzato le risorse cognitive disponibili nel sistema in modo funzionalmente adeguato al tipo di asserzione. È *epistemicamente negligente* quando il processo ha sistematicamente escluso o silenziato informazioni rilevanti che erano accessibili nel sistema. È *strutturalmente fraudolenta* quando la struttura del processo di convergenza è stata deliberatamente progettata per produrre un certo tipo di output indipendentemente dalle evidenze disponibili.

Queste categorie hanno conseguenze dirette per il dibattito contemporaneo sulla disinformazione istituzionale. La post-verità non è soltanto un fenomeno di singoli agenti che mentono consapevolmente: è, molto spesso, il risultato di strutture di convergenza che producono sistematicamente output epistemicamente degradati. L'analisi della disinformazione istituzionale richiede quindi l'analisi delle strutture di convergenza che la producono, non solo l'identificazione di individui malintenzionati.

III. Il discorso politico come sistema distribuito: conseguenze per la filosofia politica del linguaggio

La filosofia politica ha un interesse specifico negli atti linguistici perché la legittimità del potere politico è mediata, in una democrazia liberale, dal linguaggio: le leggi sono testi, la rappresentanza è costituita da atti dichiarativi, il consenso si esprime attraverso procedure verbali (voto, petizione, deliberazione pubblica). Una pragmatica del distribuito ha quindi conseguenze dirette per la teoria della legittimità politica.

La tradizione deliberativista — da Habermas a Rawls a Cohen — fonda la legittimità politica sulla qualità del processo comunicativo che produce le decisioni collettive. Le condizioni dell'ideale discorsivo habermasiano — inclusività, parità di accesso, assenza di coercizione, orientamento al consenso argomentato — definiscono uno standard normativo rispetto al quale le pratiche comunicative politiche reali possono essere valutate. Ma questo ideale presuppone che i partecipanti alla deliberazione siano soggetti unitari capaci di formare e esprimere intenzioni comunicative genuine: il discorso ideale di Habermas è un dialogo tra parlanti griceani.

La pragmatica del distribuito mette sotto pressione questo fondamento. Non perché la deliberazione non abbia valore — ha enormi ragioni per

avercelo — ma perché la descrizione dei soggetti della deliberazione come agenti con intenzioni unitarie è empiricamente inadeguata. Le parti che partecipano al discorso politico reale sono sistemi distribuiti: partiti politici con correnti in conflitto, gruppi di interesse con componenti eterogenei, movimenti sociali con leadership policefale, corpi intermedi attraversati da tensioni interne. Le loro posizioni discorsive sono il prodotto di processi di convergenza interna, e la qualità della deliberazione pubblica dipende non solo dalla qualità degli scambi tra le parti ma anche dalla qualità dei processi di convergenza interni a ciascuna.

Questo suggerisce un ampliamento del campo normativo della teoria deliberativista: le condizioni di legittimità del discorso politico non riguardano solo le caratteristiche degli scambi pubblici tra le parti, ma anche le caratteristiche interne dei processi attraverso cui ciascuna parte forma la propria posizione discorsiva. Una democrazia genuinamente deliberativa richiede non solo che il dibattito pubblico sia libero e inclusivo, ma anche che i sistemi distribuiti che vi partecipano abbiano strutture di convergenza interne democraticamente sane — che le intenzioni che portano ai tavoli della deliberazione pubblica non siano il prodotto di processi interni che silenzierebbero le voci dissidenti, concentrerebbero il potere decisionale in pochi o escluderebbero informazioni rilevanti.

Questa conseguenza — che la legittimità democratica si estende alle strutture interne di governance dei soggetti politici collettivi — è importante e relativamente inesplorata. La democrazia interna dei partiti, la trasparenza dei processi deliberativi delle grandi organizzazioni, la governance delle istituzioni sovranazionali: queste questioni sono solitamente trattate come problemi di organizzazione interna senza conseguenze dirette sulla legittimità del discorso pubblico. La pragmatica del distribuito suggerisce invece che abbiano una rilevanza costitutiva per la qualità dell'agire linguistico politico.

IV. Performatività istituzionale e critica ideologica

La teorizzazione della performatività, sviluppata autonomamente dalla teoria femminista e queer — da Butler in poi — ha prodotto una riflessione sulla forza degli atti linguistici che è complementare alla tradizione Austin-Searle pur nascendo da premesse diverse. La

performatività nel senso butleriano è la capacità di certi atti linguistici di *costituire* i soggetti e le identità che pretendono di descrivere: "è una bambina" non constata ma produce una certa posizione nel sistema di genere.

La pragmatica del distribuito può arricchire questa tradizione fornendole un'analisi più precisa dei meccanismi attraverso cui gli atti performativi istituzionali producono i loro effetti. La critica ideologica del linguaggio — da Marx al marxismo culturale, dalla critica del discorso di Foucault agli studi critici del discorso di van Dijk — ha sempre insistito sul fatto che certi atti linguistici producono effetti di naturalizzazione e di subordinazione che non sono riducibili alle intenzioni dei parlanti individuali. Il sistema capitalistico produce certi tipi di enunciazione non perché i capitalisti individuali intendano produrli, ma perché le strutture di convergenza del sistema economico tendono a generarli sistematicamente.

Questo punto — che la critica ideologica del linguaggio ha sempre intuito ma raramente formalizzato — trova nella pragmatica del distribuito una base teorica più solida. Gli effetti ideologici del linguaggio istituzionale sono spiegabili come conseguenze delle strutture di convergenza che producono certi tipi di atti linguistici, indipendentemente dalle intenzioni dei singoli contribuenti. Non è necessario postulare un soggetto con intenzioni manipolative consapevoli per spiegare perché certi sistemi producano sistematicamente discorsi che normalizzano gerarchie di potere, escludono certe voci, rendono invisibili certe realtà. È sufficiente analizzare le strutture di convergenza — chi vi partecipa, con quale peso, secondo quali procedure, con accesso a quali informazioni — per comprendere perché certe posizioni vengano sistematicamente sovrarappresentate nell'output comunicativo del sistema.

Questa analisi ha conseguenze pratiche per il design istituzionale. Se si vuole che un sistema distribuito produca atti linguistici epistemicamente più equilibrati e meno distorsivi, è necessario intervenire sulle sue strutture di convergenza: ampliare la varietà dei contribuenti, equalizzare i pesi, diversificare le fonti di informazione, introdurre procedure di controllo che consentano alle voci minoritarie di incidere sull'output. La riforma istituzionale, in questa prospettiva, è anche — e forse primariamente — una riforma delle strutture di convergenza che determinano quale tipo di linguaggio l'istituzione produce.

V. Intelligenza artificiale e responsabilità linguistica: il caso limite

La proliferazione dei sistemi di intelligenza artificiale generativa pone la teoria sviluppata in questa monografia di fronte al suo caso limite più urgente e, in certi sensi, più rivelatore. I grandi modelli linguistici producono atti comunicativi — asserzioni, consigli, diagnosi, previsioni — che hanno effetti pratici reali sulle persone e sulle organizzazioni che li ricevono. La questione di chi sia responsabile di questi atti, e in base a quale principio, è diventata una delle questioni più pressanti del diritto e dell'etica tecnologica contemporanei.

La teoria classica della pragmatica non offre risorse per affrontare questa questione perché i modelli linguistici non hanno intenzioni nel senso psicologico richiesto dalla teoria. La risposta giuridica più immediata — che la responsabilità ricade sull'azienda che ha sviluppato il sistema — è una soluzione di ripiego che trasferisce il problema senza risolverlo: come si determina la responsabilità dell'azienda per un output prodotto da un sistema il cui comportamento non è completamente prevedibile nemmeno dai suoi sviluppatori?

La pragmatica del distribuito offre un'analisi più precisa. Un modello linguistico è, da questo punto di vista, un sistema di convergenza straordinariamente complesso: le sue risposte sono il prodotto di processi distribuiti attraverso miliardi di parametri, ciascuno dei quali porta il contributo statistico di una porzione del corpus su cui il modello è stato addestrato. La "intenzione" dell'atto linguistico prodotto dal modello non è localizzabile in nessun componente: è il risultato emergente di un processo di convergenza globale.

Tuttavia, la struttura di questo processo non è completamente opaca alla valutazione normativa. Si può distinguere tra la responsabilità relativa al *design della funzione di convergenza* (le scelte di architettura, di addestramento, di allineamento che determinano le caratteristiche statistiche degli output del sistema), la responsabilità relativa al *contesto di deployment* (le condizioni in cui il sistema è messo in uso, i tipi di query a cui è esposto, le informazioni contestuali che influenzano le sue risposte), e la responsabilità relativa all'*uso* (le scelte dell'utente che guidano la conversazione e che co-producono l'atto linguistico finale).

Questa tripartizione non è esaustiva, ma è teoricamente più adeguata di soluzioni monistiche che collocano tutta la responsabilità in un unico punto. In particolare, essa consente di articolare la responsabilità dei produttori di sistemi AI non solo in termini di difetti di progettazione (come nel diritto dei prodotti) ma in termini di scelte di design della struttura di convergenza — che tipi di output il sistema è stato addestrato a produrre, su quali corpus, con quali criteri di allineamento, con quali meccanismi di controllo. Questa prospettiva sposta il campo della responsabilità dal dominio degli incidenti imprevedibili a quello delle scelte progettuali deliberate, che è il dominio in cui la responsabilità ha più senso essere imputata.

VI. La comunicazione istituzionale come pratica di costruzione della convergenza

Le conseguenze pratiche della pragmatica distribuita per la comunicazione istituzionale vanno oltre la questione della responsabilità. Includono una riconcettualizzazione di cosa sia la *buona* comunicazione istituzionale e di quali siano i suoi requisiti normativi.

Nella prospettiva tradizionale, la comunicazione istituzionale è buona quando è chiara, accurata, coerente e tempestiva — quando trasmette efficacemente le intenzioni e le posizioni dell'istituzione ai suoi interlocutori. In questa prospettiva, i problemi di comunicazione istituzionale sono essenzialmente problemi tecnici: difetti di redazione, scelte lessicali inadeguate, strutture argomentative difettose.

Nella prospettiva della pragmatica distribuita, la comunicazione istituzionale è buona quando riflette adeguatamente la complessità e la genuinità del processo di convergenza che l'ha prodotta. Un comunicato che presenta come unanime e coerente una posizione che è stata raggiunta attraverso conflitti interni non risolti non è solo strategicamente rischioso (perché le fratture interne prima o poi emergeranno); è epistemicamente disonesto rispetto agli interlocutori, che ricevono un'immagine distorta del sistema con cui si relazionano. Una comunicazione istituzionale genuinamente onesta dovrebbe — nei casi in cui è appropriato — rendere visibile la struttura distributiva del processo che l'ha prodotta: le incertezze residue, le posizioni minoritarie, i limiti epistemici del sistema.

Questo è un ideale regolativo, non una descrizione di come le organizzazioni comunicano abitualmente. Ma come ideale regolativo ha un contenuto normativo preciso e non banale: indica che la trasparenza comunicativa delle istituzioni non riguarda solo il contenuto di ciò che dicono ma anche la struttura dei processi attraverso cui lo decidono. La governance trasparente è, da questo punto di vista, anche una condizione di buona pragmatica istituzionale.

VII. La filosofia politica del linguaggio: oltre il monopolio della parola pubblica

L'ultima conseguenza teorica che merita esplorazione riguarda la distribuzione del potere linguistico nella sfera pubblica. La tradizione liberal-democratica ha sempre riconosciuto l'importanza della libertà di espressione come condizione della democrazia, ma ha tipicamente concettualizzato questa libertà come una proprietà di individui: il diritto del singolo a esprimere le proprie opinioni senza censura. Anche quando si è riconosciuto che i sistemi di comunicazione di massa creano asimmetrie di potere comunicativo — che la libertà di espressione senza eguaglianza di accesso ai mezzi di comunicazione è libertà formale senza sostanza — la risposta teorica è rimasta prevalentemente individualistica: garantire a ogni individuo un accesso equo alla parola pubblica.

La pragmatica del distribuito suggerisce un frame diverso. Il problema non è solo quello dell'accesso individuale al discorso pubblico, ma quello delle strutture di convergenza attraverso cui i soggetti collettivi — partiti, organizzazioni, istituzioni — producono le loro posizioni pubbliche. Una democrazia in cui ogni individuo ha formalmente diritto di parola, ma in cui le strutture di convergenza dei grandi attori collettivi sono opache, gerarchiche e non democraticamente controllate, è una democrazia comunicativamente zoppa. Il potere reale nel discorso pubblico non appartiene agli individui che esprimono opinioni personali: appartiene ai sistemi distribuiti che producono atti linguistici con forza istituzionale.

Da questo punto di vista, la *democratizzazione delle strutture di convergenza* — rendere più democratiche le procedure attraverso cui le grandi organizzazioni e istituzioni formano le loro posizioni pubbliche — è un progetto politico di filosofia del linguaggio. Implica norme sulla trasparenza dei processi deliberativi interni, requisiti di pluralismo nella composizione dei sistemi che producono atti linguistici ad alto impatto

pubblico, garanzie di protezione per le voci dissenzienti all'interno dei sistemi organizzativi, diritti di accesso all'informazione che consentano ai cittadini di valutare la struttura dei processi di convergenza che producono il discorso istituzionale a cui sono esposti.

Nessuna di queste misure è rivoluzionaria in sé: molte esistono già, in forma embrionale, nelle normative sulla governance societaria, sulla trasparenza amministrativa, sulla responsabilità dei media. Ciò che la pragmatica distribuita aggiunge è una giustificazione teorica più precisa per queste misure — non solo come strumenti di lotta alla corruzione o alla concentrazione del potere economico, ma come condizioni di possibilità di una pragmatica pubblica genuinamente democratica. E con questa giustificazione, suggerisce anche quali siano le lacune più rilevanti: i sistemi distribuiti che producono atti linguistici ad alto impatto pubblico senza essere soggetti ad alcun requisito di trasparenza dei propri processi di convergenza sono un problema non solo politico ma specificamente linguistico-pragmatico.

Casi empirici della convergenza distribuita

I. Premessa metodologica: il caso empirico come banco di prova teorico

Una teoria filosofica del linguaggio non è tale in senso pieno finché rimane confinata nell'ordine dell'argomentazione astratta. La sua forza dipende anche dalla capacità di illuminare casi concreti che resisterebbero a descrizioni alternative: casi che la teoria classica non può spiegare, o può spiegare solo a costo di distorsioni significative, e che invece trovano nella pragmatica dell'intenzione distribuita una resa descrittiva più adeguata. Questa parte si propone di esaminare cinque aree empiriche in cui la struttura distribuita degli atti linguistici è particolarmente evidente e le conseguenze pratiche particolarmente rilevanti.

I casi scelti non sono campionati in modo statisticamente rappresentativo — un'operazione che richiederebbe strumenti empirici che esulano da questa monografia — ma in modo teoricamente mirato: ciascuno mette sotto pressione una componente specifica del quadro teorico sviluppato nella terza parte, consentendo di verificarne la tenuta e di identificarne i limiti. Il caso delle agenzie di rating interroga la

funzione di convergenza in condizioni di asimmetria informativa estrema e di conflitto di interesse strutturale. Il caso della legislazione europea interroga la dimensione temporale della convergenza e il problema dell'ambiguità intenzionale. I rapporti dell'IPCC mettono alla prova la distinzione tra condizioni strutturali e sostanziali di felicità in un contesto di produzione scientifica collettiva. Le crisi reputazionali aziendali illuminano il meccanismo dell'assunzione di stabilità funzionale in condizioni avverse. Infine, le dichiarazioni di guerra e i trattati di pace — atti linguistici di forza estrema — consentono di esaminare il funzionamento della forza illocutoria procedurale nelle circostanze più gravide di conseguenze.

II. Le agenzie di rating come sistemi linguistici distribuiti: 2007–2009

Le agenzie di rating del credito — Moody's, Standard & Poor's, Fitch — occupano una posizione peculiare nel sistema finanziario globale: producono atti linguistici (i ratings) che hanno forza performativa diretta sui mercati, ma lo fanno attraverso strutture organizzative la cui distribuzione interna è sistematicamente oscurata dalla forma unitaria dell'output finale. Un rating è, dal punto di vista della pragmatica distribuita, un atto illocutorio ibrido: ha la forma di un'asserzione (si afferma qualcosa circa la solvibilità di un emittente) ma la forza di una dichiarazione (il solo atto di emettere il rating modifica la realtà che pretende di descrivere, influenzando i costi di finanziamento dell'emittente e le scelte degli investitori).

La crisi finanziaria del 2007–2009 ha reso visibile in modo drammatico la struttura distribuita di questi atti linguistici e i guasti che questa struttura può produrre. I ratings AAA attribuiti a strumenti strutturati basati su mutui subprime — i famosi CDO e RMBS che avrebbero perso la quasi totalità del loro valore entro il 2009 — erano il prodotto di processi di convergenza interna alle agenzie caratterizzati da asimmetrie strutturali di informazione, pressioni commerciali sistemiche e meccanismi di validazione reciproca tra analisti.

La ricostruzione documentata dalle indagini del Congresso americano e dalla Financial Crisis Inquiry Commission consente di identificare almeno quattro sotto-sistemi in competizione all'interno del processo di convergenza che produceva i ratings:

Il primo sotto-sistema era quello degli analisti quantitativi, responsabili della costruzione dei modelli matematici su cui si basava la valutazione. Questi analisti operavano con assunzioni statistiche — in particolare, basse correlazioni storiche tra i tassi di default dei mutui in diverse aree geografiche — che erano note essere problematiche in condizioni di stress sistemico, ma che erano difficilmente modificabili senza compromettere la capacità competitiva dell'agenzia in un mercato in cui la concorrenza si giocava anche sulla generosità delle valutazioni.

Il secondo sotto-sistema era quello commerciale, i cui incentivi erano allineati con la produzione di ratings favorevoli agli emittenti che pagavano le commissioni. Il modello dell'emittente pagante — issuer pays model — creava una pressione strutturale verso ratings più generosi che non richiedeva nessun atto di corruzione esplicita: era sufficiente che gli analisti commerciali selezionassero, tra le metodologie disponibili, quelle che producevano risultati più favorevoli agli emittenti.

Il terzo sotto-sistema era quello del risk management interno all'agenzia, che avrebbe dovuto svolgere una funzione di controllo sulle valutazioni degli analisti di credito. Ma le evidenze documentate mostrano che questo sotto-sistema aveva una posizione istituzionalmente subordinata rispetto al processo produttivo principale: i suoi allarmi venivano sistematicamente attenuati o ignorati nel processo di convergenza che portava all'output finale.

Il quarto sotto-sistema era quello della conformità regolamentare, anch'esso teoricamente indipendente ma praticamente integrato con la cultura produttiva dell'organizzazione.

Il rating finale era il prodotto della convergenza tra questi quattro sotto-sistemi, con pesi molto diversi: il sotto-sistema commerciale esercitava un'influenza sproporzionata non attraverso istruzioni esplicite ma attraverso la selezione dei framework metodologici, la determinazione dei parametri dei modelli e la struttura degli incentivi degli analisti.

Questa analisi consente di applicare la distinzione tra condizioni strutturali e sostanziali di felicità con risultati significativi. I ratings erano strutturalmente validi: erano stati prodotti secondo le procedure interne delle agenzie, da analisti con le qualificazioni previste, secondo metodologie formalmente documentate. Ma erano sostanzialmente fallaci in un senso preciso: il processo di convergenza aveva

sistematicamente escluso o sottoponderato le informazioni disponibili nel sistema circa i limiti delle assunzioni modellistiche e i rischi di correlazione in condizioni di stress. Non si trattava di menzogna individuale — la stragrande maggioranza degli analisti credeva nelle proprie valutazioni — ma di un guasto strutturale nella funzione di convergenza che produceva outputs epistemicamente degradati indipendentemente dalle intenzioni dei singoli partecipanti.

La categoria della responsabilità strutturale, introdotta nella quarta parte, trova qui un'applicazione diretta. Le agenzie erano responsabili non perché i loro dipendenti avessero inteso ingannare (sebbene in alcuni casi documentati ciò si sia verificato), ma perché avevano progettato e mantenuto strutture di convergenza che erano funzionalmente orientate a produrre un certo tipo di output indipendentemente dalla qualità epistemica dei dati disponibili. La responsabilità per design della convergenza — che si propone come terzo livello del quadro responsabilità nella pragmatica distribuita — è precisamente il tipo di responsabilità che i tribunali e i regolatori hanno avuto difficoltà a imputare usando le categorie tradizionali, e che invece trova una collocazione teorica precisa nel quadro qui proposto.

Un aspetto ulteriore che merita attenzione è il meccanismo dell'implicatura in questo contesto. L'assunzione di stabilità funzionale — quella che sostituisce l'assunzione di cooperatività griceana nella pragmatica distribuita — prevedeva che le agenzie di rating avessero un interesse funzionale nel mantenere la coerenza tra i propri ratings e la qualità creditizia effettiva degli emittenti, perché dalla credibilità dei loro giudizi dipendeva la loro sopravvivenza nel mercato. Questa assunzione si è rivelata, nel periodo 2003–2007, un fondamento fragile: le agenzie avevano sviluppato una posizione oligopolistica che attenuava le conseguenze reputazionali degli errori sistematici, e il loro interesse funzionale immediato era diventato quello di massimizzare le commissioni piuttosto che la qualità epistemica degli output. Quando l'assunzione di stabilità funzionale viene meno, il meccanismo delle implicature si incrina: gli investitori che avevano inferito, dall'emissione di un rating AAA, che l'agenzia disponesse di evidenze adeguate a supportare quel giudizio, avevano tratto un'implicatura che il sistema non poteva sostenere.

III. La legislazione europea come atto linguistico distribuito: il caso delle direttive

La produzione normativa dell'Unione Europea costituisce probabilmente il caso più complesso e analiticamente ricco di atto linguistico distribuito nella storia moderna. Una direttiva europea — si prenda come esempio paradigmatico la Direttiva sul Mercato Interno dell'Elettricità (Direttiva 2019/944/CE) o la più recente AI Act (Regolamento 2024/1689/UE) — è il prodotto di un processo di convergenza che coinvolge: la Commissione Europea (che detiene il monopolio dell'iniziativa legislativa), il Parlamento Europeo (che rappresenta i cittadini degli stati membri attraverso gruppi parlamentari transnazionali), il Consiglio dell'Unione Europea (che rappresenta i governi degli stati membri), i gruppi di lavoro tecnici e le agenzie dell'UE, le organizzazioni di lobbying delle industrie interessate, i consulenti giuridici delle istituzioni, e — in ultima analisi — le corti nazionali e la Corte di Giustizia dell'UE che interpretano e applicano il testo.

Il processo legislativo ordinario — la cosiddetta procedura legislativa ordinaria o codecisione — prevede fino a tre "letture" in cui il testo viene modificato da Parlamento e Consiglio, seguite dai trilogues informali in cui rappresentanti delle tre istituzioni negoziano il testo finale in sessioni che non sono pubbliche e i cui verbali non vengono sistematicamente pubblicati. Il testo finale di una direttiva è, letteralmente, il sedimento di questo processo: contiene le tracce delle negoziazioni, dei compromessi, delle ambiguità deliberatamente introdotte per consentire a posizioni incompatibili di convergere su un testo comune.

Questa struttura produce un fenomeno che nel quadro teorico della pragmatica distribuita si può denominare ambiguità strutturale deliberata: l'atto linguistico finale contiene ambiguità che non sono accidentali (il prodotto di una redazione imprecisa) né semantiche (il prodotto di un linguaggio intrinsecamente vago) ma intenzionali a livello di sistema, nel senso che nessun sub-sistema avrebbe potuto convergere su un testo univoco senza rinunciare a posizioni che considerava irrinunciabili.

Il caso dell'AI Act è istruttivo. Il testo finale — approvato con larga maggioranza dal Parlamento Europeo nel marzo 2024 e formalmente in vigore dall'agosto dello stesso anno — contiene definizioni di "sistema di IA ad alto rischio" che sono state deliberatamente mantenute aperte per consentire la convergenza tra la posizione del Parlamento (favorevole a definizioni più ampie e a obblighi più stringenti) e quella del Consiglio

(più attenta alle preoccupazioni delle imprese e alla competitività europea). Il risultato è un testo che le imprese, i regolatori e i giuristi interpretano in modo significativamente divergente, e la cui applicazione uniforme richiederà probabilmente anni di giurisprudenza della CGUE.

Dal punto di vista della teoria classica degli atti linguistici, questo è un fallimento normativo: un atto linguistico con condizioni di soddisfazione insufficientemente determinate. Dal punto di vista della pragmatica distribuita, è un esito strutturalmente prevedibile e teoricamente comprensibile: la funzione di convergenza ha prodotto un output che massimizza l'accordo formale tra i sub-sistemi al costo di ridurre la determinatezza del contenuto. La tensione tra completezza della convergenza (tutti i sub-sistemi accettano il testo) e determinatezza dell'output (il testo ha condizioni di soddisfazione chiare) è una tensione strutturale dei sistemi distribuiti di grandi dimensioni, non una disfunzione contingente.

La dimensione temporale della convergenza è qui particolarmente rilevante. Il processo di trasposizione delle direttive negli ordinamenti nazionali — in cui ciascuno stato membro produce atti linguistici nazionali (leggi, decreti, regolamenti) che devono essere conformi al testo europeo ma che riflettono anche le strutture di convergenza interne al sistema politico-legislativo nazionale — introduce un secondo livello di distribuzione che moltiplica le possibilità di divergenza interpretativa. La "stessa" direttiva può produrre obblighi significativamente diversi in stati diversi, non perché i legislatori nazionali stiano violando il diritto europeo, ma perché le strutture di convergenza nazionali selezionano, tra le interpretazioni plausibili del testo ambiguo, quelle più compatibili con le loro preferenze preesistenti.

Questo fenomeno — che il diritto europeo chiama "goldplating" quando gli stati aggiungono requisiti più stringenti di quelli della direttiva, ma che la pragmatica distribuita descrive più precisamente come ri-convergenza locale di un atto linguistico la cui convergenza originale era incompleta — è una delle fonti principali di frammentazione del mercato interno europeo. La sua spiegazione in termini di strutture di convergenza distribuite suggerisce anche i rimedi più efficaci: non solo meccanismi di enforcement più robusti, ma interventi sulla struttura del processo legislativo europeo che riducano l'incentivo a produrre ambiguità deliberate come meccanismo di convergenza.

IV. I rapporti dell'IPCC: scienza collettiva come pragmatica del distribuito

I rapporti dell'Intergovernmental Panel on Climate Change costituiscono un caso paradigmatico di atto linguistico distribuito in cui la struttura della convergenza è esplicitamente documentata e le sue conseguenze per la qualità epistemica dell'output sono state oggetto di riflessione sistematica da parte degli stessi produttori. Pochi esempi nel panorama contemporaneo rendono così visibile la tensione tra le diverse componenti della funzione di convergenza.

Un rapporto dell'IPCC — si prenda il Sesto Rapporto di Valutazione (AR6), il cui Summary for Policymakers (SPM) è stato approvato nel 2021–2022 — è il prodotto di tre sotto-sistemi in convergenza strutturalmente distinti. Il primo è quello degli autori scientifici: centinaia di scienziati provenienti da decine di paesi, coordinati in working groups tematici, che sintetizzano migliaia di studi peer-reviewed in valutazioni integrate della letteratura. Il secondo è quello del processo di revisione: esperti indipendenti e rappresentanti governativi revisionano le bozze dei rapporti tecnici in cicli successivi, producendo commenti che gli autori sono tenuti a considerare (e a cui devono rispondere per iscritto se scelgono di non incorporarli). Il terzo è quello intergovernativo: il Summary for Policymakers viene approvato "riga per riga" dai rappresentanti dei governi dei 195 paesi membri, in sessioni plenarie che durano spesso più giorni della durata prevista e in cui ogni parola è oggetto di negoziazione.

Questa struttura produce una distinzione di fatto — anche se non sempre di diritto — tra il corpo principale dei rapporti tecnici (Technical Reports) e i Summaries for Policymakers. I Technical Reports sono prodotti dalla convergenza tra autori scientifici e revisori, e hanno come criterio di validità la coerenza con la letteratura scientifica disponibile. Gli SPM sono prodotti dalla convergenza ulteriore tra il prodotto tecnico e le preferenze dei governi, e hanno come criterio di validità aggiuntivo l'accettabilità per una comunità di stati con interessi molto diversi rispetto all'azione climatica.

Questa distinzione è pregnante per la teoria delle condizioni di felicità. L'SPM è strutturalmente valido quando viene approvato all'unanimità dai governi seguendo le procedure IPCC. Ma le sue condizioni sostanziali di felicità — la corrispondenza tra il contenuto dell'asserzione e le credenze giustificate disponibili nel sistema — dipendono da una struttura di

convergenza che ha introdotto pressioni sistematiche verso l'attenuazione dei messaggi più allarmanti e l'ampliamento delle formulazioni che consentono interpretazioni più o meno stringenti.

Gli studi sulle differenze sistematiche tra i Technical Reports e i corrispondenti SPM hanno documentato un fenomeno noto come "erring on the side of least drama": una tendenza sistematica dei documenti sottoposti a convergenza intergovernativa a sottostimare i rischi rispetto a quanto indicato dalla letteratura scientifica sottostante. Questo fenomeno non è attribuibile a malafede dei rappresentanti governativi o degli autori scientifici: è il prodotto strutturale di una funzione di convergenza in cui i partecipanti con le preferenze più conservative esercitano un potere di veto de facto (poiché l'approvazione richiede il consenso).

Il caso IPCC illumina anche il meccanismo dell'implicatura in contesti distribuiti. Quando l'IPCC afferma, in uno SPM, che è "estremamente probabile" che il riscaldamento globale osservato sia causato dall'attività umana, l'implicatura attesa è che questa formulazione rispecchi fedelmente il grado di certezza espresso dalla comunità scientifica nei Technical Reports. Ma se la funzione di convergenza intergovernativa tende sistematicamente ad attenuare le formulazioni rispetto ai documenti tecnici, questa implicatura potrebbe essere fuorviante: il grado di certezza effettivamente espresso dalla letteratura scientifica potrebbe essere superiore a quello implicato dall'uso di una certa categoria di probabilità nell'SPM.

La risposta dell'IPCC a questa tensione — la decisione di mantenere la distinzione tra Technical Reports e SPM, e di rendere pubblicamente disponibili entrambi i documenti — può essere interpretata come un tentativo di costruire trasparenza sulla struttura del processo di convergenza: il lettore che ha accesso a entrambi i documenti può inferire dove si sono prodotte le divergenze e valutare autonomamente il grado di attenuazione introdotto dalla convergenza intergovernativa. Questo è un esempio di buona prassi comunicativa nella pragmatica distribuita: rendere visibile la struttura della convergenza come strumento di onestà epistemica nei confronti degli interlocutori.

V. Le crisi reputazionali aziendali come fallimenti della convergenza

Le grandi crisi reputazionali aziendali degli ultimi decenni — dal caso Enron alla crisi Volkswagen sulle emissioni diesel, dall'epidemia di salmonella di Chipotle alla gestione comunicativa della crisi Boeing 737 MAX — offrono un terreno empirico eccezionalmente ricco per la pragmatica distribuita, perché rendono visibili, ex post, le strutture di convergenza che avevano prodotto gli atti linguistici problematici e i meccanismi attraverso cui queste strutture avevano sistematicamente generato output epistemicamente o eticamente degradati.

Il caso Volkswagen è particolarmente istruttivo per la teoria della responsabilità distribuita. La "defeat device" — il software che modificava il comportamento del motore durante i cicli di test per le emissioni — era nota a un numero significativo di ingegneri e dirigenti all'interno dell'azienda. Eppure il management di vertice ha sostenuto, inizialmente e con alcune prove di supporto, di non essere stato informato. Questa affermazione ha sollevato una domanda teoricamente urgente: come è possibile che un'azienda di quelle dimensioni abbia prodotto per anni comunicazioni pubbliche — rapporti di sostenibilità, pubblicità comparative sulle emissioni, comunicazioni ai regolatori — che affermavano conformità normativa, senza che i livelli superiori della struttura di convergenza conoscessero la falsità di queste affermazioni?

La risposta della pragmatica distribuita è che questo è esattamente il tipo di output che strutture di convergenza mal progettate possono produrre sistematicamente. Una struttura organizzativa in cui le informazioni tecniche negative non hanno canali appropriati per emergere verso i livelli decisionali superiori, in cui gli incentivi premiano il raggiungimento degli obiettivi di performance indipendentemente dai mezzi, e in cui la cultura organizzativa scoraggia la comunicazione di problemi "insolubili", è una struttura di convergenza che produrrà atti linguistici epistemicamente mal fondati anche in assenza di intenzioni fraudolente esplicite ai livelli superiori della gerarchia.

La responsabilità per design della convergenza — il terzo livello del quadro proposto nella quarta parte — si applica con precisione: chi ha progettato e mantenuto una struttura organizzativa che rendeva sistematicamente impossibile la comunicazione verticale delle informazioni tecniche critiche è responsabile degli atti linguistici pubblici che quella struttura ha prodotto, anche se non conosceva il contenuto specifico di quei problemi.

Il caso Boeing 737 MAX offre un ulteriore livello di analisi. I comunicati aziendali sulla sicurezza del velivolo dopo i due incidenti del 2018 e del 2019 — che causarono la morte di 346 persone — erano il prodotto di una struttura di convergenza in cui il sotto-sistema ingegneristico della sicurezza aveva già prodotto valutazioni interne allarmanti, ma in cui il sotto-sistema della comunicazione esterna operava con accesso limitato a queste informazioni e con una funzione di convergenza orientata principalmente alla gestione della crisi reputazionale piuttosto che alla comunicazione accurata dei rischi. Il risultato fu una serie di atti linguistici pubblici — dichiarazioni dei CEO, comunicati stampa, comunicazioni ai regolatori — che erano strutturalmente validi (prodotti dai soggetti formalmente competenti seguendo le procedure previste) ma sostanzialmente fallaci (non rispecchiavano le credenze disponibili nel sistema circa i rischi effettivi dell'aeromobile).

VI. Le dichiarazioni di guerra e i trattati di pace: atti linguistici di forza estrema

Le dichiarazioni di guerra e i trattati di pace costituiscono, dal punto di vista della pragmatica distribuita, il caso limite della forza illocutoria procedurale. Si tratta di atti linguistici la cui forza è massimale — modificano radicalmente la realtà giuridica, politica e spesso fisica di intere popolazioni — e la cui distribuzione è talmente evidente da rendere paradossale qualunque tentativo di ricondurli a un'intenzione unitaria.

La dichiarazione di guerra degli Stati Uniti all'Iraq nel 1991 — tecnicamente, l'autorizzazione al Presidente all'uso della forza militare, approvata dal Congresso — è il prodotto di un processo legislativo che coinvolse centinaia di legislatori con posizioni molto diverse, lobbying intensissimi, pressioni dell'esecutivo, considerazioni di politica estera complessa e calcoli elettorali individuali. Il voto finale — 52 a 47 al Senato, margine storico — non rappresenta un'intenzione condivisa nel senso griceano: rappresenta la convergenza procedurale di posizioni profondamente divergenti su un testo formale che autorizzava un'azione con conseguenze che nessuno dei votanti poteva prevedere con certezza.

Eppure questo atto linguistico aveva forza performativa piena: scatenò una guerra, modificò le relazioni internazionali per decenni, produsse obblighi giuridici, morali e politici per un'intera nazione. La forza non

derivava dall'intenzione comune dei partecipanti al voto ma dalla posizione che l'atto occupava nella struttura istituzionale: era stato prodotto dal organo competente, seguendo le procedure costituzionali, con il quorum necessario. La forza illocutoria procedurale opera qui nella sua forma più pura.

I trattati di pace offrono una complessità ancora maggiore. Il Trattato di Versailles del 1919 — uno degli atti linguistici più conseguenziali del XX secolo — fu il prodotto di una negoziazione tra le potenze vincitrici della Prima Guerra Mondiale caratterizzata da profondi conflitti interni a ciascuna delegazione nazionale (il Presidente Wilson era in aperto conflitto con il Congresso americano; Lloyd George negoziava sotto pressione dell'opinione pubblica britannica; Clemenceau bilanciava obiettivi di sicurezza francese con pressioni degli alleati). Il testo finale era, come ha documentato la storiografia in modo ormai definitivo, il sedimento di compromessi instabili tra posizioni incompatibili, in cui l'ambiguità deliberata era la forma della convergenza su punti cruciali come le riparazioni di guerra e i confini territoriali.

Le conseguenze di questa struttura di convergenza — l'instabilità del testo, le interpretazioni divergenti tra le potenze firmatarie, il crescente risentimento tedesco verso clausole percepite come punitive — sono state al centro del dibattito storico e politico per un secolo. Dal punto di vista della pragmatica distribuita, queste conseguenze erano strutturalmente prevedibili: un atto linguistico prodotto da una convergenza provvisoria e conflittuale ha una forza normativa intrinsecamente instabile, che dipende dal continuo mantenimento dell'equilibrio tra le forze che lo hanno prodotto. Quando questo equilibrio si rompe — come si è rotto rapidamente nel caso di Versailles con il ritiro americano dalla Società delle Nazioni e l'ascesa del nazionalismo tedesco — la forza dell'atto si incrina, e con essa la realtà politica che pretendeva di costituire.

Questa analisi ha implicazioni normative per il design dei processi di produzione degli atti linguistici internazionali. La stabilità di un trattato non dipende solo dalla qualità del suo contenuto proposizionale — da quanto giuste o equilibrate siano le sue disposizioni — ma dalla struttura della convergenza che lo ha prodotto: dalla robustezza degli accordi interni alle parti firmatarie, dalla profondità del consenso raggiunto sulle questioni più controverse, dalla presenza di meccanismi di revisione che consentano di adattare il testo al mutare delle condizioni senza invalidare la convergenza di base. La pragmatica distribuita suggerisce che la

durabilità degli accordi internazionali — un obiettivo cruciale per la stabilità del sistema internazionale — è una proprietà della struttura della convergenza più che del contenuto specifico dell'accordo.

VII. Sintesi: pattern ricorrenti nei casi empirici

L'esame di questi cinque casi empirici — pur nella loro eterogeneità — rivela pattern ricorrenti che confermano e precisano le ipotesi teoriche sviluppate nelle parti precedenti.

Il primo pattern è quello della dissociazione sistematica tra validità strutturale e fondatezza sostanziale. In tutti i casi esaminati, gli atti linguistici problematici erano strutturalmente validi: prodotti dai soggetti formalmente competenti, seguendo le procedure previste, nelle forme richieste. Il loro fallimento — epistemico, normativo o pratico — era un fallimento sostanziale, riconducibile a strutture di convergenza che escludevano o sottoponderavano informazioni rilevanti disponibili nel sistema. Questo pattern suggerisce che i criteri di valutazione degli atti linguistici distribuiti debbano essere espansi: la validità formale è necessaria ma non sufficiente.

Il secondo pattern è quello della responsabilità diffusa e del deficit imputativo. In tutti i casi, la responsabilità per gli atti linguistici problematici era distribuita tra numerosi agenti nessuno dei quali era integralmente responsabile dell'output finale. Le categorie giuridiche tradizionali — dolo, colpa, concorso di persone — si sono rivelate inadeguate a catturare questa distribuzione, producendo esiti processuali spesso insoddisfacenti per le vittime dei danni. La responsabilità strutturale e la responsabilità per design della convergenza offrono categorie più precise, ma la loro applicazione pratica richiede sviluppi dottrinali e normativi ancora in gran parte da costruire.

Il terzo pattern è quello dell'ambiguità intenzionale come meccanismo di convergenza. In situazioni di forte conflitto tra sotto-sistemi — la negoziazione legislativa europea, la produzione dei ratings in condizioni di conflitto di interesse, la diplomazia internazionale — l'ambiguità deliberata del testo finale è spesso il prezzo della convergenza: il modo in cui posizioni incompatibili trovano un accordo rinunciando alla determinatezza del contenuto. Questa ambiguità ha costi sistemici che si manifestano nelle fasi successive (interpretazione, applicazione,

enforcement) ma ha un beneficio immediato (rendere possibile la convergenza). Il design delle strutture di convergenza dovrebbe cercare di ridurre il ricorso a questo meccanismo, sviluppando procedure alternative che consentano la convergenza senza sacrificare la determinatezza.

Il quarto pattern — forse il più importante per la teoria — è quello della forza illocutoria procedurale come risorsa normativa autonoma. La forza degli atti linguistici istituzionali non dipende dalla qualità della convergenza che li ha prodotti ma dalla posizione che occupano nella struttura istituzionale. Questo è al tempo stesso una caratteristica funzionale essenziale delle istituzioni (garantisce la certezza e la prevedibilità del diritto) e una fonte di rischio sistemico (può conferire forza normativa ad atti epistemicamente o eticamente degradati). La tensione tra queste due dimensioni è irrisolvibile in astratto; può essere gestita attraverso il design di strutture di convergenza che massimizzano la qualità sostanziale degli output riducendo la probabilità di dissociazione tra validità formale e fondatezza epistemica.

Formalizzazione della funzione di convergenza

I. La necessità di un apparato formale

La descrizione intuitiva della funzione di convergenza proposta nella terza parte — come processo che trasforma la molteplicità di vettori intenzionali distribuiti in un output comunicativo dotato di forza illocutoria determinata — ha il merito di rendere visibile la struttura del fenomeno e di consentire confronti tra casi. Ma rimane, nella sua forma attuale, troppo metaforica per supportare le inferenze normative e predittive che una teoria matura deve essere in grado di produrre. In questa parte si tenterà una formalizzazione più precisa, attingendo a tre apparati teorici: la teoria dei giochi cooperativi, la teoria delle reti e la logica deontica multi-agente.

Va detto con franchezza che questa formalizzazione è ancora parziale e programmatica: il suo scopo non è presentare un sistema assiomatico completo ma identificare le strutture formali rilevanti, mostrare la loro adeguatezza descrittiva rispetto ai fenomeni di interesse, e indicare le direzioni in cui una formalizzazione più completa dovrebbe procedere.

L'impresa di costruire una pragmatica formale del distribuito è un progetto di ricerca che coinvolgerà necesariamente competenze interdisciplinari — filosofi del linguaggio, teorici dei giochi, informatici, giuristi — e che questa monografia non può che abbozzare.

Un avvertimento metodologico è necessario. La formalizzazione non è un fine in sé: il rischio di ogni formalizzazione in filosofia è quello di catturare precisamente qualcosa di leggermente diverso da ciò che interessa, sacrificando la rilevanza per la rigorosità. Nel caso della pragmatica distribuita, il rischio specifico è di costruire modelli formali che descrivono bene sistemi ideali di deliberazione razionale ma perdono di vista la dimensione normativa — ciò che rende un processo di convergenza buono o cattivo, legittimo o illegittimo. Il criterio di adeguatezza dei modelli formali proposti non è solo la coerenza interna ma la capacità di illuminare questa dimensione normativa.

II. Il framework della teoria dei giochi cooperativi

La teoria dei giochi cooperativi — sviluppata da von Neumann e Morgenstern e approfondita da Shapley, Aumann e altri — studia come agenti con preferenze potenzialmente divergenti possano formare coalizioni e distribuire i benefici della cooperazione. Questa teoria offre strumenti formali potenzialmente utili per modellare la funzione di convergenza, con alcune qualificazioni importanti.

Definizione 1 (Gioco di convergenza). Sia $N = {1, 2, \ldots, n}$ l'insieme degli agenti che partecipano al processo di convergenza. Sia Σ lo spazio degli output comunicativi possibili (l'insieme di tutti gli atti linguistici che il sistema potrebbe produrre). Per ogni agente $i \in N$, sia $u_i : \Sigma \to \mathbb{R}$ la funzione di utilità che esprime la preferenza dell'agente sull'output. Un gioco di convergenza $G = (N, \Sigma, {u_i}_{i \in N}, \mathcal{P})$ è definito dai suddetti elementi più un insieme $\mathcal{P}$ di procedure ammissibili — le regole che il processo deve seguire per produrre un output strutturalmente valido.

Questa definizione cattura gli elementi essenziali: la molteplicità degli agenti, la loro eterogeneità nelle preferenze, lo spazio degli output possibili, e i vincoli procedurali che distinguono la convergenza da un semplice mercato delle intenzioni. La distinzione tra convergenza strutturalmente valida (prodotta seguendo le procedure di

$\mathcal{P}$) e convergenza ottimale secondo qualche criterio sostanziale è già incorporata nella struttura del gioco.

Definizione 2 (Nucleo del gioco di convergenza). Il nucleo $C(G)$ del gioco di convergenza è l'insieme degli output $\sigma \in \Sigma$ tali che nessuna coalizione propria $S \subsetneq N$ avrebbe potuto, attraverso le procedure ammissibili per la coalizione S, produrre un output che tutti i membri di S preferissero a σ. Un output nel nucleo è stabile nel senso che nessun sottoinsieme degli agenti ha incentivo a deviare unilateralmente dal processo di convergenza globale.

La nozione di nucleo è rilevante per la teoria della forza illocutoria procedurale: un output stabile — nel nucleo del gioco di convergenza — ha maggiori probabilità di generare una convergenza duratura e di mantenere la propria forza normativa nel tempo. Un output fuori dal nucleo è in equilibrio instabile: esistono coalizioni che preferirebbero un diverso output e che, se trovassero i meccanismi appropriati, agirebbero per modificarlo. La fragilità del Trattato di Versailles — discussa nella parte precedente — può essere formalizzata come un output al di fuori del nucleo del gioco di convergenza internazionale: esisteva una coalizione (Germania + URSS) che aveva strumenti per destabilizzarlo e che aveva preferenze incompatibili con il suo mantenimento.

Proposizione 1 (Esistenza del nucleo e struttura della convergenza). Il nucleo $C(G)$ è non vuoto se e solo se il gioco di convergenza è bilanciato nel senso di Bondareva-Shapley. La bilanciabilità è una condizione sull'allineamento delle preferenze degli agenti: in termini intuitivi, il nucleo è non vuoto quando non esiste nessun modo per le coalizioni di bloccarsi vicendevolmente in modo da rendere impossibile qualunque accordo stabile.

Questa proposizione ha un'interpretazione normativa immediata: le strutture di convergenza che producono sistemi bilanciati — in cui le procedure garantiscono che le preferenze degli agenti possano trovare un punto di equilibrio stabile — producono atti linguistici con forza normativa più robusta di quelle che producono sistemi sbilanciati, in cui l'output è sempre il frutto di una coalizione vincente instabile.

Il problema della dipendenza dal sentiero. La teoria dei giochi cooperativi standard assume che le coalizioni si formino simultaneamente e che l'output sia determinato dall'equilibrio finale, indipendentemente dall'ordine in cui le coalizioni si sono formate.

Questa assunzione è inadeguata per i processi di convergenza reali, in cui l'ordine dei contributi conta enormemente: chi parla per primo, quale proposta viene messa sul tavolo inizialmente, quale compromesso viene raggiunto nella fase di agenda-setting — tutti questi elementi influenzano sistematicamente l'output finale.

Per catturare questa dimensione occorre estendere il framework con una struttura di ordinamento temporale. Sia $T = {t_0, t_1, \ldots, t_K}$ una sequenza di stadi temporali del processo di convergenza. In ciascuno stadio t_k, un sotto-insieme di agenti $N_k \subseteq N$ produce contributi al processo che modificano lo stato del processo s_k secondo una funzione di transizione $\delta: S \times A \to S$, dove S è lo spazio degli stati possibili e A è l'insieme delle azioni disponibili. L'output finale σ è determinato dallo stato finale s_K.

Questo modello dinamico cattura la dipendenza dal sentiero e consente di modellare fenomeni come il path-locking — la situazione in cui l'output finale è fortemente vincolato dalle scelte fatte nelle fasi iniziali del processo, anche se queste non erano ottimali — e il strategic agenda-setting — la manipolazione dell'ordine del giorno per influenzare l'output attraverso la strutturazione del processo piuttosto che attraverso la qualità degli argomenti.

III. Modelli di rete per l'intenzione distribuita

La teoria dei giochi cooperativi descrive la struttura degli incentivi e degli equilibri del processo di convergenza, ma non cattura la topologia delle interazioni — il fatto che gli agenti non interagiscono tutti con tutti, ma attraverso reti strutturate in cui la posizione di ciascun agente determina la sua influenza sull'output finale. La teoria delle reti complesse offre strumenti complementari che consentono di modellare questa dimensione.

Definizione 3 (Rete di convergenza). Una rete di convergenza $\mathcal{N} = (V, E, w, \phi)$ è un grafo pesato orientato in cui:

- $V = {v_1, \ldots, v_n}$ è l'insieme dei nodi, corrispondenti agli agenti nel processo;

- $E \subseteq V \times V$ è l'insieme degli archi, che rappresentano canali di comunicazione e influenza tra agenti;
- $w: E \to \mathbb{R}^+$ è una funzione di peso che quantifica l'intensità dell'influenza trasmessa lungo ciascun arco;
- $\phi: V \to \Delta(\Sigma)$ è una funzione che assegna a ciascun nodo una distribuzione di probabilità sullo spazio degli output possibili, rappresentando le preferenze dell'agente.

La struttura della rete — la sua topologia — determina come le preferenze dei nodi si propagano attraverso il sistema per determinare l'output finale. Le misure di centralità standard della teoria delle reti — grado, betweenness, PageRank — acquisiscono una interpretazione pragmatica precisa in questo contesto.

Centralità e peso pragmatico. La centralità di betweenness di un nodo v in una rete di convergenza misura la frequenza con cui v si trova sul percorso più breve tra coppie di altri nodi: è una misura del ruolo di intermediazione che v svolge nel sistema. In una rete di convergenza, un nodo con alta betweenness è un intermediario strategico: controlla i flussi di informazione e influenza tra agenti che non sono direttamente connessi. Questo ruolo non è necessariamente correlato alla posizione gerarchica formale: un funzionario di rango medio che coordina la comunicazione tra dipartimenti può avere una betweenness molto alta e quindi un'influenza sull'output della convergenza sproporzionata rispetto al suo ruolo formale.

La centralità di PageRank — adattata al contesto della convergenza — misura l'influenza di un nodo tenendo conto non solo del numero di connessioni ma del peso di chi lo connette: un nodo influenzato da molti nodi centrali è più centrale di uno influenzato da molti nodi periferici. Questa misura cattura meglio del grado semplice la struttura effettiva del potere nelle reti di convergenza istituzionali.

Dinamiche di propagazione e sincronizzazione. In una rete di convergenza, le preferenze dei nodi non sono fisse ma si modificano attraverso l'interazione: gli agenti aggiornano le proprie preferenze in risposta alle preferenze espresse dagli agenti con cui sono connessi. Questo processo — che può essere modellato attraverso dinamiche di opinione su reti — determina se e come il sistema converge verso un output condiviso.

Si consideri il modello DeGroot di aggiornamento delle credenze su reti: ciascun agente i aggiorna la propria credenza $x_i^{(t)}$ allo stadio t come media ponderata delle credenze dei propri vicini al tempo precedente:

$$x_i^{(t+1)} = \sum_{j \in N(i)} w_{ij} \cdot x_j^{(t)}$$

dove w_{ij} è il peso che l'agente i attribuisce all'agente j e $N(i)$ è l'insieme dei vicini di i nella rete. Sotto condizioni appropriate sulla struttura della rete (connettività, aperiodicità), questo processo converge a un consenso stazionario la cui forma dipende dalla struttura dei pesi e dalla distribuzione iniziale delle credenze.

Questo modello cattura un fenomeno importante per la pragmatica distribuita: la polarizzazione può emergere come esito stabile del processo di aggiornamento quando la rete è strutturata in modo tale da isolare gruppi di nodi con preferenze simili, riducendo l'influenza degli agenti "ponte" tra cluster diversi. Le camere d'eco nei social media — reti di convergenza distribuita informale — sono un caso paradigmatico: la topologia della rete produce sistematicamente atti linguistici (posizioni pubbliche, narrative condivise) che rispecchiano le preferenze del cluster più che dell'intera comunità.

Reti di convergenza istituzionale: strutture tipiche. Le reti di convergenza delle istituzioni formali tendono ad avere strutture topologiche diverse da quelle delle reti sociali informali. Le strutture gerarchiche — organigrammi piramidali — producono reti in cui pochi nodi apicali hanno gradi molto elevati e la betweenness è concentrata: l'output della convergenza è dominato dalle preferenze dei nodi centrali, e la struttura di convergenza tende a riprodurre le preferenze del vertice con attenuazioni marginali dai livelli inferiori.

Le strutture collegiali — consigli, commissioni, assemblee — producono reti più dense con distribuzione di centralità più uniforme: l'output della convergenza è più sensibile alle preferenze della maggioranza, ma anche più vulnerabile a manipolazioni attraverso il controllo dell'agenda (chi decide cosa viene messo ai voti) e delle procedure di votazione (votazione palese vs. segreta, soglie di maggioranza, ordine delle opzioni).

Le strutture reticolari — reti di istituzioni che interagiscono orizzontalmente senza gerarchia formale, come molti sistemi di

governance internazionale — producono reti sparse con molti nodi di betweenness elevata: l'output della convergenza è fortemente dipendente dall'attività degli intermediari strategici e dalle coalizioni che si formano tra nodi non direttamente connessi.

IV. Logica deontica multi-agente per la pragmatica distribuita

La terza componente del framework formale è la logica deontica — la logica delle norme, degli obblighi e dei permessi — estesa al contesto multi-agente. Questa estensione è necessaria per formalizzare la dimensione normativa della funzione di convergenza: le condizioni di felicità degli atti linguistici distribuiti non possono essere espresse solo in termini di preferenze e utilità, ma richiedono un linguaggio che distingua tra ciò che il sistema deve fare, può fare, e non può fare nell'esercizio delle proprie funzioni comunicative.

Il linguaggio DMAL. Si propone un frammento di logica deontica multi-agente (DMAL — Deontic Multi-Agent Logic) adeguato per la pragmatica distribuita. Il linguaggio DMAL ha:

- Una sintassi proposizionale di base per descrivere stati del mondo (compresi gli stati degli atti linguistici: "l'atto α è stato compiuto", "la condizione C è soddisfatta")
- Operatori modali per agenti individuali: $K_i \phi$ ("l'agente i sa che ϕ"), $B_i \phi$ ("l'agente i crede che ϕ"), $I_i \phi$ ("l'agente i intende che ϕ")
- Operatori deontici per agenti individuali: $O_i \phi$ ("l'agente i ha l'obbligo che ϕ"), $P_i \phi$ ("l'agente i ha il permesso che ϕ"), $F_i \phi$ ("è vietato all'agente i che ϕ")
- Operatori collettivi: $K_C \phi$ ("la coalizione C sa collettivamente che ϕ"), $O_C \phi$ ("la coalizione C ha l'obbligo collettivo che ϕ")
- Operatori per atti linguistici: $Assert(i, \phi)$ ("l'agente i asserisce ϕ"), $Promise(i, \phi)$ ("l'agente i promette ϕ"), e così via per i tipi illocutori rilevanti.
- Operatori per la convergenza: $Conv(C, \alpha)$ ("la coalizione C converge sull'atto linguistico α"), $Valid_P(C,

\alpha)$ ("l'atto α è strutturalmente valido per la coalizione C secondo la procedura P")

Formalizzazione delle condizioni di felicità distribuite. Le condizioni di felicità degli atti linguistici distribuiti possono essere espresse in DMAL con precisione maggiore di quanto fosse possibile nella descrizione informale della terza parte.

Per una *promessa istituzionale* — ad esempio, una delibera con cui un consiglio si impegna a compiere un'azione futura — le condizioni strutturali di felicità sono:

$$Valid_P(N, Promise(N, \phi)) \iff \left(\exists P \in \mathcal{P}\right) \left[P\text{-processo}(N) \land \text{quorum}(N, P)\right]$$

Le condizioni sostanziali richiedono che il sistema abbia accesso alle credenze rilevanti:

$$SubstFelicit(Promise(N, \phi)) \iff K_N \text{can}(N, \phi) \land K_N \text{context}(\phi)$$

dove K_N è conoscenza collettiva della coalizione N, $\text{can}(N, \phi)$ esprime la capacità del sistema di mantenere l'impegno ϕ, e $\text{context}(\phi)$ le condizioni contestuali rilevanti per la sua esecuzione.

La dissociazione tra condizioni strutturali e sostanziali — discussa nella terza e quinta parte — è ora formalmente precisa: un atto linguistico distribuito α è strutturalmente valido ma sostanzialmente fallace quando $Valid_P(N, \alpha) \land \neg SubstFelicit(\alpha)$.

Teorema di responsabilità distribuita (formulazione formale). Un risultato importante che il framework DMAL consente di esprimere con precisione è la seguente condizione necessaria per la responsabilità di un agente individuale i nell'ambito di un processo di convergenza collettivo:

$$Resp(i, \alpha) \Rightarrow \left[K_i \text{contribution}(i, \alpha) \lor B_i \text{contribution}(i, \alpha)\right] \land \left[P_i \text{prevent}(\neg \alpha) \lor O_i \text{inform}(N \setminus \{i\}, \neg SubstFelicit(\alpha))\right]$$

In termini informali: l'agente i è responsabile dell'atto linguistico distribuito α se (a) sapeva o credeva di contribuire alla produzione di α, e (b) aveva o il potere di prevenire α o l'obbligo di informare il sistema della mancata soddisfazione delle condizioni sostanziali di felicità.

Questa formalizzazione cattura la struttura della responsabilità epistemica distribuita introdotta nella quarta parte e la rende operazionalizzabile: per attribuire responsabilità a un agente, occorre verificare se aveva accesso alle informazioni rilevanti circa le condizioni sostanziali dell'atto e se aveva strumenti (formali o informali) per incidere sul processo di convergenza.

V. Misure di qualità della convergenza

Un framework formale per la pragmatica distribuita deve includere misure che consentano di confrontare la qualità di diverse strutture di convergenza — di dire che una convergenza è epistemicamente più adeguata, più equa, più robusta di un'altra. Si propongono tre misure distinte, corrispondenti a tre dimensioni normative della convergenza.

Misura 1: Adeguatezza epistemica. Sia $\mathcal{E}(N)$ l'insieme di tutte le informazioni epistemicamente rilevanti disponibili nel sistema — la conoscenza distribuita tra i nodi della rete. Sia $\mathcal{E}(Conv(N, \alpha))$ l'insieme delle informazioni che hanno effettivamente influenzato il processo di convergenza che ha prodotto α. L'adeguatezza epistemica $\epsilon(\alpha)$ della convergenza è:

$$\epsilon(\alpha) = \frac{|\mathcal{E}(Conv(N, \alpha)) \cap \mathcal{E}(N)|}{|\mathcal{E}(N)|}$$

cioè la proporzione delle informazioni rilevanti disponibili nel sistema che hanno effettivamente influenzato la convergenza. Una convergenza è epistemicamente adeguata nella misura in cui mobilizza le risorse cognitive del sistema piuttosto che sistematicamente escluderle.

Questa misura formalizza la distinzione tra strutture di convergenza epistemicamente virtuose (che massimizzano ϵ) e strutture epistemicamente viziose (che producono ϵ basso escludendo sistematicamente informazioni rilevanti). Il caso Volkswagen discusso nella quinta parte si caratterizza per un basso ϵ nella

convergenza comunicativa esterna: le informazioni relative alla defeat device erano disponibili nel sistema ma non hanno influenzato la produzione dei comunicati pubblici.

Misura 2: Equità della convergenza. La funzione di convergenza distribuisce il peso pragmatico tra gli agenti in modo potenzialmente asimmetrico: le preferenze di certi agenti pesano più di altre nel determinare l'output finale. Una misura naturale di questa asimmetria è l'indice di Gini applicato alla distribuzione dei pesi pragmatici:

$$Gini(\mathcal{F}) = \frac{\sum_{i=1}^n \sum_{j=1}^n |w_i - w_j|}{2n \sum_{i=1}^n w_i}$$

dove w_i è il peso pragmatico dell'agente i — la misura in cui le sue preferenze determinano l'output finale. Un indice di Gini uguale a zero indica una convergenza perfettamente equa (ogni agente pesa ugualmente); un indice prossimo a uno indica una convergenza fortemente concentrata (un singolo agente determina l'output).

Questa misura ha un'interpretazione normativa pluralista: strutture di convergenza con indice di Gini basso sono, ceteris paribus, preferibili da un punto di vista democratico. Ma il ceteris paribus è cruciale: strutture di convergenza con basso Gini e bassa adeguatezza epistemica — sistemi in cui molti agenti contribuiscono ugualmente con informazioni scarse o irrilevanti — possono produrre output peggiori di strutture più asimmetriche con alta adeguatezza epistemica.

Misura 3: Robustezza della convergenza. La robustezza di una convergenza misura la stabilità dell'output rispetto a perturbazioni del sistema: quanto cambierebbe l'output se un agente modificasse le proprie preferenze, se un nuovo agente venisse aggiunto al processo, o se le procedure venissero modificate?

Formalmente, la robustezza $\rho(\alpha)$ può essere definita come:

$$\rho(\alpha) = 1 - \mathbb{E}\left[\frac{d(\alpha, \alpha')}{d_{max}}\right]$$

dove α' è l'output prodotto da una piccola perturbazione del sistema, $d(\cdot, \cdot)$ è una metrica sullo spazio degli output comunicativi, e l'aspettativa è presa rispetto alla distribuzione delle perturbazioni possibili. Una convergenza robusta $(\rho \approx 1)$ produce output stabili rispetto a piccoli cambiamenti nel sistema; una

convergenza fragile $(\rho \approx 0)$ produce output sensibili a perturbazioni minime.

La robustezza corrisponde, nella teoria degli atti linguistici distribuiti, alla durabilità della forza normativa discussa nella sesta sezione della terza parte. Atti linguistici prodotti da convergenze robuste — come le norme costituzionali nei sistemi politici stabili — mantengono la propria forza anche al mutare delle condizioni; atti prodotti da convergenze fragili — come i trattati internazionali basati su coalizioni instabili — perdono forza al cambiare degli equilibri di potere.

VI. Il teorema di Arrow e i limiti della convergenza

Una delle asimmetrie più illuminanti tra la teoria dei giochi cooperativi e la pragmatica distribuita riguarda le implicazioni del teorema di impossibilità di Arrow. Questo risultato — uno dei più profondi della teoria della scelta sociale — dimostra che non esiste nessuna funzione di aggregazione delle preferenze che soddisfi simultaneamente un insieme di condizioni ragionevoli: unanimità (se tutti gli agenti preferiscono x a y, la funzione di aggregazione preferisce x a y), indipendenza delle alternative irrilevanti (il ranking relativo di x e y dipende solo dai ranking individuali di x e y, non da quello di una terza alternativa z), non-dittatura (nessun agente singolo determina sempre l'output), e dominio illimitato (la funzione funziona per qualsiasi combinazione di preferenze individuali).

Tradotto nella terminologia della pragmatica distribuita, il teorema di Arrow afferma che non esiste nessuna funzione di convergenza F che sia al tempo stesso epistemicamente adeguata (rispetti l'unanimità), indipendente dal framing (rispetti l'indipendenza delle alternative irrilevanti), democratica (non-dittatura), e universalmente applicabile (dominio illimitato).

Questa impossibilità ha conseguenze profonde per la teoria normativa delle strutture di convergenza. Significa che nessuna struttura di convergenza può soddisfare simultaneamente tutti i criteri normativi che si potrebbero ragionevolmente chiedere: ogni struttura reale rappresenta un compromesso tra questi criteri, e la scelta del compromesso è essa stessa una scelta normativa che richiede giustificazione.

In particolare, il teorema di Arrow suggerisce che le strutture di convergenza reali devono necessariamente violare almeno una delle condizioni desiderabili:

- Le strutture gerarchiche violano la non-dittatura: il vertice della gerarchia esercita un'influenza determinante sull'output.
- Le strutture deliberative basate su argomentazione razionale tendono a violare l'indipendenza delle alternative irrilevanti: il framing della discussione influenza sistematicamente l'output.
- Le strutture di voto a maggioranza in domini multi-dimensionali producono ciclicità (paradosso di Condorcet) che violano la transitività delle preferenze collettive.
- Le strutture basate su expertise selettiva violano l'unanimità: le opinioni di agenti non esperti vengono sistematicamente sottopesate.

La pragmatica normativa del distribuito non può quindi aspirare a identificare la struttura di convergenza ottimale in senso assoluto: deve invece articolare i trade-off rilevanti e i criteri per scegliere tra essi a seconda del contesto e degli obiettivi dell'atto linguistico da produrre. Una struttura di convergenza appropriata per produrre una sentenza giudiziaria (che valorizza la non-dittatura procedurale e l'expertise legale) è diversa da quella appropriata per produrre una politica pubblica (che valorizza l'inclusività e la rappresentanza degli interessi) o per produrre una valutazione scientifica (che valorizza l'expertise e l'adeguatezza epistemica).

VII. Algoritmi di convergenza e loro proprietà normative

Un'ultima componente del framework formale riguarda gli algoritmi attraverso cui la convergenza viene realizzata — i meccanismi concreti attraverso cui i sistemi distribuiti producono output comunicativi. Diversi algoritmi di convergenza hanno proprietà normative diverse, e la scelta dell'algoritmo non è mai neutra rispetto all'output prodotto.

Algoritmo 1: Aggregazione lineare. L'algoritmo più semplice assegna pesi $\lambda_i \geq 0$ (con $\sum_i \lambda_i = 1$) a ciascun agente e produce l'output come media ponderata delle preferenze:

$$\sigma^* = \arg\max_{\sigma \in \Sigma} \sum_{i=1}^n \lambda_i u_i(\sigma)$$

Questo algoritmo è normativamente semplice: produce un output che massimizza la somma pesata delle utilità, con proprietà note di trattabilità computazionale. Il suo difetto principale è che ignora la struttura di rete delle interazioni: tratta ogni agente come se interagisse direttamente con tutti gli altri, senza considerare la mediazione degli intermediari.

Algoritmo 2: Deliberazione iterativa. Il processo di deliberazione iterativa — tipico dei sistemi deliberativi formali come le assemblee parlamentari — procede per stadi: in ciascuno stadio, gli agenti aggiornano le proprie preferenze alla luce degli argomenti presentati dagli altri, e il processo termina quando si raggiunge un punto fisso (consenso) o una condizione di uscita (voto).

Questo algoritmo ha la proprietà, dimostrata da Habermas nell'ambito della sua teoria dell'azione comunicativa, di tendere verso soluzioni basate sulla "forza del migliore argomento" in condizioni ideali. Ma richiede condizioni strutturali molto stringenti (parità di accesso, assenza di coercizione, tempo sufficiente) che raramente si verificano nella pratica, e la sua convergenza non è garantita in presenza di preferenze cicliche.

Algoritmo 3: Mediazione e procedure di Delphi. Le procedure di tipo Delphi — originariamente sviluppate per la previsione tecnologica ma applicabili a qualunque processo di convergenza collettivo — prevedono cicli iterativi di raccolta delle preferenze, anonimizzazione, feedback aggregato agli agenti e revisione delle preferenze. L'anonimizzazione riduce le distorsioni dovute a fattori di status e autorità formale; il feedback aggregato consente agli agenti di aggiornare le proprie posizioni alla luce delle preferenze degli altri senza essere vincolati a difendere posizioni pubblicamente assunte.

Questo algoritmo ha proprietà normative interessanti: l'anonimizzazione aumenta l'adeguatezza epistemica (riducendo il peso degli status gerarchici) e l'equità (riducendo la concentrazione del peso pragmatico). Il suo limite principale è la difficoltà di applicarlo in contesti in cui la responsabilità dell'output deve essere attribuibile a individui specifici — contesti in cui l'anonimizzazione è quindi incompatibile con le strutture di responsabilità esistenti.

Convergenza algoritmica e sistemi AI. I sistemi di intelligenza artificiale generativa costituiscono un caso limite degli algoritmi di convergenza: il loro output è determinato da un processo computazionale che può essere descritto come una forma di convergenza su distribuzioni di probabilità attraverso miliardi di parametri. Dal punto di vista della teoria qui proposta, la domanda normativa rilevante è: quali sono le proprietà dell'algoritmo di convergenza implementato da un modello linguistico rispetto alle dimensioni di adeguatezza epistemica, equità e robustezza?

La risposta è complessa e dipende in modo critico dalle scelte fatte nelle fasi di pre-addestramento (quale corpus, con quali pesi), fine-tuning (quale tipo di feedback umano, da quale pool di annotatori) e allineamento (quali vincoli comportamentali, con quale struttura di preferenze). Queste scelte determinano la struttura della funzione di convergenza implementata dal sistema e le sue proprietà normative — un punto che la quinta parte della monografia aveva identificato come cruciale per la teoria della responsabilità nell'IA, e che il framework formale qui sviluppato consente di precisare: la responsabilità dei produttori di sistemi AI è precisamente la responsabilità per le scelte normative incorporate nell'algoritmo di convergenza.

Epistemologia del testimone distribuito

I. Il problema epistemico della voce collettiva

Quando un sistema distribuito produce un'asserzione — "il prodotto è sicuro", "l'economia cresce", "il clima si sta riscaldando a un ritmo senza precedenti" — sorge immediatamente una domanda filosofica che la tradizione epistemologica classica non ha mai affrontato con sufficiente rigore: chi è il soggetto epistemico di questa asserzione? A chi appartiene la credenza che pretende di esprimere? E cosa significa, per un sistema distribuito, sapere qualcosa?

L'epistemologia moderna — da Cartesio a Gettier, dalla tradizione della giustificazione internalista alle epistemologie della testimonianza — ha costruito il suo apparato concettuale intorno al soggetto individuale. La conoscenza è una relazione tra un soggetto e una proposizione: il soggetto crede che p, p è vera, e la credenza è giustificata in modo

appropriato. Anche le epistemologie sociali più sofisticate — da Fricker a Goldman, da Coady a Zagzebski — che hanno allargato il campo d'indagine alla conoscenza prodotta attraverso l'interazione sociale (testimonianza, expertise, deliberazione collettiva) rimangono fondamentalmente ancorate al soggetto individuale come unità di analisi: studiano come i soggetti individuali acquisiscano conoscenza attraverso gli altri, non come i sistemi distribuiti possiedano e producano conoscenza propria.

La pragmatica dell'intenzione distribuita richiede un'epistemologia che faccia il passo ulteriore: che concepisca la conoscenza distribuita non come aggregato di conoscenze individuali ma come proprietà emergente di sistemi cognitivi collettivi. Questo progetto — che può essere chiamato epistemologia del testimone distribuito — ha radici in alcune tradizioni filosofiche marginali (la filosofia del senso comune di Reid, certe letture di Wittgenstein, l'epistemologia evolutiva) ma non ha ancora una sistemazione coerente. Questa parte ne abbozza i fondamenti, con attenzione particolare alle conseguenze per la teoria degli atti linguistici distribuiti.

II. Tre modelli di conoscenza distribuita

Prima di sviluppare una teoria positiva, è utile distinguere tre modelli di come la conoscenza può essere distribuita in un sistema multi-agente. Questi modelli non sono esclusivi — i sistemi reali combinano elementi di ciascuno — ma rappresentano strutture logicamente distinte con proprietà epistemiche molto diverse.

Il modello partitivo. Nel modello partitivo, la conoscenza del sistema è la somma delle conoscenze degli agenti individuali: ciascun agente possiede una parte dell'informazione totale, e il sistema "sa" l'unione di queste parti. Questo modello è quello implicito nei sistemi informativi distribuiti in senso tecnico: database replicati, archivi distribuiti, reti di sensori. La proprietà fondamentale del modello partitivo è l'additività: la conoscenza del sistema aumenta monotonicamente al crescere della conoscenza degli agenti individuali.

Il limite del modello partitivo per la pragmatica distribuita è la sua incapacità di catturare i fenomeni di emergenza epistemica: situazioni in cui il sistema "sa" qualcosa che nessun agente individuale sa, non perché

l'informazione sia semplicemente distribuita ma perché emerge dall'interazione tra le parti. Fenomeni come le bolle speculative — in cui il mercato nel suo complesso "valuta" gli asset in modo che nessun singolo investitore potrebbe individualmente giustificare — o la saggezza delle folle — in cui le stime aggregate di gruppi non esperti superano le stime degli esperti individuali — non sono spiegabili dal modello partitivo.

Il modello emergente. Nel modello emergente, la conoscenza del sistema non è riducibile alla somma delle conoscenze degli agenti: emerge dalle interazioni tra agenti attraverso processi che non sono trasparenti ai partecipanti. La "conoscenza" di un mercato sui prezzi relativi, la "saggezza" di una giuria sulla colpevolezza di un imputato, il "consenso scientifico" su una questione empirica — in tutti questi casi, c'è qualcosa che il sistema "sa" nel senso che il suo comportamento aggregato riflette, e che nessun agente individuale potrebbe articolare completamente.

Il modello emergente è epistemicamente più ricco ma anche più problematico: i processi di emergenza non sono necessariamente virtuosi. Le bolle speculative sono un fenomeno emergente in cui il mercato "sa" qualcosa di falso — attribuisce valore a asset privi di fondamentali adeguati — per un periodo di tempo che può essere molto lungo. Le camere d'eco nei social media producono credenze emergenti sistematicamente distorte. L'emergenza epistemica non garantisce la verità; può produrre sia conoscenza genuina che errore collettivo.

Il modello normativo. Nel modello normativo, la conoscenza del sistema è determinata non dalle credenze psicologiche degli agenti ma dalla posizione che certi output occupano nella struttura istituzionale. Il "sapere" di un tribunale — la verità giudiziale di un fatto — non dipende da ciò che i giudici credono individualmente ma dalla procedura attraverso cui è stata prodotta: prove ammissibili, standard probatori, garanzie processuali. Analogamente, il "sapere" di un'agenzia regolatoria — che un farmaco è sicuro ed efficace — dipende dalla procedura di approvazione seguita, non dalle credenze dei singoli valutatori.

Il modello normativo è fondamentale per la pragmatica distribuita perché cattura la forza illocutoria procedurale degli atti linguistici istituzionali: la verità di una sentenza o di un'approvazione regolatoria non è la verità epistemica ordinaria ma una verità normativamente costituita. Questo non significa che sia arbitraria — le procedure sono progettate per

produrre approssimazioni della verità epistemica — ma che ha una struttura logica diversa che richiede un'analisi epistemologica specifica.

III. La testimonianza distribuita: struttura e problemi

La tradizione epistemologica della testimonianza — sviluppata in modo sistematico a partire dai lavori di Coady (1992), con contributi fondamentali di Fricker, Goldberg, Lackey e altri — studia come i soggetti individuali acquisiscano conoscenza attraverso le affermazioni di altri. Le domande fondamentali di questa tradizione riguardano la giustificazione della credenza testimoniale: quando siamo giustificati a credere a ciò che ci viene detto? Quali sono le condizioni che trasformano una credenza acquisita per testimonianza in conoscenza?

La pragmatica distribuita richiede di estendere queste domande in una direzione diversa: non come i soggetti individuali acquisiscano conoscenza attraverso sistemi distribuiti, ma come i sistemi distribuiti producano atti di testimonianza con forza epistemica propria. Quando un'autorità sanitaria afferma che un vaccino è sicuro, quando un'agenzia di rating attribuisce un giudizio di solvibilità, quando una corte dichiara un fatto accertato — questi non sono atti di testimonianza individuale ma atti di testimonianza distribuita: il sistema distribuito funge da testimone collettivo di una realtà che pretende di descrivere.

La struttura della testimonianza distribuita è analoga a quella dell'atto linguistico distribuito in generale, ma con una componente aggiuntiva cruciale: la credibilità. Un testimone individuale è credibile nella misura in cui ha accesso all'informazione rilevante (competenza) e non ha motivo di distorcerla (sincerità). Un sistema distribuito che produce testimonianza è credibile nella misura in cui le sue strutture di convergenza garantiscono adeguatezza epistemica (mobilizzazione delle informazioni rilevanti) e assenza di distorsioni sistematiche (neutralità della funzione di convergenza rispetto agli interessi dei partecipanti).

Il problema dell'ingiustizia epistemica distribuita. Miranda Fricker ha introdotto il concetto di ingiustizia epistemica per descrivere situazioni in cui qualcuno subisce un torto nella propria capacità di agente epistemico: la testimonianza di certe persone è sistematicamente sottovalutata (ingiustizia testimoniale) o certe persone mancano degli strumenti interpretativi per dare senso alle proprie esperienze (ingiustizia

ermeneutica). Questi fenomeni, che Fricker analizza principalmente a livello delle interazioni interpersonali, hanno analoghi sistemici nelle strutture di convergenza dei sistemi distribuiti.

L'ingiustizia testimoniale distribuita si produce quando le strutture di convergenza di un sistema assegnano sistematicamente pesi epistemici bassi alle voci di certi tipi di agenti, non per ragioni legate alla loro competenza epistemica ma per ragioni di status, identità o posizione gerarchica. Un'organizzazione in cui le segnalazioni di rischio provenienti dai livelli operativi vengono sistematicamente filtrate e attenuate prima di raggiungere i livelli decisionali produce atti linguistici che riflettono un'ingiustizia testimoniale distribuita: chi ha accesso diretto all'informazione rilevante ha un peso pragmatico minimo nella funzione di convergenza.

L'ingiustizia ermeneutica distribuita si produce quando le strutture concettuali disponibili nel sistema non consentono di articolare e trasmettere certe categorie di esperienza che sono rilevanti per la produzione degli atti linguistici del sistema. Le organizzazioni che hanno strutturato il proprio sistema di reporting interno intorno a categorie di rischio che non catturano certi tipi di minacce — ad esempio, i rischi sociali o ambientali a lungo termine che non entrano nei modelli finanziari standard — producono atti linguistici che non possono incorporare queste categorie di esperienza, non perché gli agenti che ne hanno esperienza siano esclusi dal processo, ma perché mancano gli strumenti interpretativi per tradurre questa esperienza in contributi al processo di convergenza.

IV. Aggregazione delle credenze e paradossi del giudizio

La logica dell'aggregazione delle credenze — lo studio formale di come le credenze individuali possano essere combinate in credenze collettive — ha prodotto risultati che sono direttamente rilevanti per la pragmatica distribuita e che gettano luce su strutture di fallimento che altrimenti rimarrebbero opache.

Il risultato più importante in questo contesto è il paradosso del docente (discursive dilemma), formalizzato da Pettit e List (2002) a partire da un'intuizione di Kornhauser e Sager (1986) nel contesto giuridico. Il paradosso può essere illustrato con un esempio semplice.

Un comitato di tre esperti deve decidere se approvare un progetto che richiede tre condizioni: P (il progetto è tecnicamente fattibile), Q (è economicamente sostenibile) e R (ha un impatto ambientale accettabile). La deliberazione produce i seguenti giudizi:

Esperto	**P**	**Q**	**R**	**$P \land Q \land R$**
E_1	Sì	Sì	No	No
E_2	Sì	No	Sì	No
E_3	No	Sì	Sì	No
Maggioranza	**Sì**	**Sì**	**Sì**	**No**

Se si vota sulle premesse (P, Q, R) separatamente, la maggioranza approva ciascuna premessa: il progetto soddisfa tutti e tre i criteri. Ma se si vota sulla conclusione ($P \land Q \land R$), la maggioranza non approva il progetto: nessun esperto ritiene che tutte e tre le condizioni siano soddisfatte simultaneamente.

Questo paradosso — che emerge in qualunque procedura di aggregazione che combini giudizi su proposizioni logicamente connesse — ha implicazioni profonde per la teoria degli atti linguistici distribuiti. Mostra che la scelta della procedura di convergenza non è neutrale rispetto all'output: votare sulle premesse o sulla conclusione produce risultati diversi, e nessuno dei due risultati è necessariamente più "corretto" dell'altro dal punto di vista della razionalità collettiva.

La scelta tra procedure basate sulle premesse e procedure basate sulla conclusione è, nella terminologia della pragmatica distribuita, una scelta sulla struttura della funzione di convergenza che ha conseguenze dirette sulla forza illocutoria dell'atto linguistico prodotto. Un sistema che adotta procedure basate sulle premesse — delibera separatamente su

ogni componente della questione prima di combinare i giudizi — produce atti linguistici con una struttura argomentativa più trasparente e analizzabile, ma è soggetto all'incongruenza del paradosso del docente. Un sistema che adotta procedure basate sulla conclusione — chiede direttamente un giudizio sull'output finale — produce atti linguistici con maggiore coerenza interna ma con minore trasparenza sul ragionamento sottostante.

Questa tensione non ha una soluzione universalmente ottimale. La scelta dipende dal tipo di atto linguistico che si intende produrre e dalle priorità normative del sistema. Nei sistemi giuridici, la tendenza prevalente — espressa nella struttura dei procedimenti collegiali — è quella di richiedere unanimità o maggioranza qualificata sulla conclusione (il dispositivo della sentenza) lasciando libertà di ragionamento individuale sulle premesse (le motivazioni, che possono divergere anche quando il dispositivo è condiviso). Questa scelta produce atti linguistici con forza illocutoria robusta (la sentenza è chiara) ma ragionamento distribuito (le motivazioni possono essere contraddittorie), che è precisamente la struttura che una teoria degli atti linguistici distribuiti deve essere in grado di descrivere.

V. Bolle epistemiche e cascate informative nei sistemi distribuiti

Una delle patologie più rilevanti delle strutture di convergenza — documentata empiricamente in una vasta letteratura in economia comportamentale, psicologia sociale e teoria delle organizzazioni — è quella della bolla epistemica: la situazione in cui un sistema distribuito converge su una credenza collettiva che si discosta sistematicamente dalla verità, non perché i singoli agenti siano irrazionali o disonesti, ma per ragioni strutturali legate alla topologia delle interazioni e agli incentivi del processo di convergenza.

Le cascate informative, teorizzate da Bikhchandani, Hirshleifer e Welch (1992), mostrano come comportamenti gregari possano emergere razionalmente da strutture di incentivi ordinarie. Si consideri una sequenza di agenti che devono prendere una decisione binaria (adottare o meno una tecnologia, investire o meno in un'azienda, sostenere o meno una posizione politica) in base alla propria informazione privata e all'osservazione delle scelte precedenti. Se i primi agenti della sequenza

convergono sulla stessa scelta per ragioni contingenti, gli agenti successivi possono trovare razionale conformarsi alla scelta prevalente anche quando la propria informazione privata suggerirebbe il contrario: il segnale aggregato dalle scelte osservate può essere più forte del segnale privato individuale.

Il risultato è una cascata: una convergenza collettiva su una credenza che può essere sistematicamente errata, e che è estremamente robusta rispetto alla revisione perché ogni nuovo agente razionale si conforma alla cascata esistente, rafforzandola. Le bolle speculative dei mercati finanziari — dalle tulipomanie del XVII secolo alla bolla delle dot-com del 2000 — sono stati interpretati come cascate informative in cui la convergenza del sistema su valutazioni errate è stata mantenuta da incentivi strutturali alla conformità.

Dal punto di vista della pragmatica distribuita, le cascate informative sono casi di convergenza epistemicamente degradata: il sistema produce atti assertivi (valutazioni di prezzo, posizioni di investimento, giudizi di qualità) la cui forza epistemica è inversamente proporzionale alla robustezza della loro fondazione informativa. Il paradosso è che queste asserzioni tendono ad avere alta robustezza nel senso della terza misura proposta nella sesta parte — l'output è stabile rispetto alle perturbazioni — proprio perché il meccanismo della cascata rende il sistema insensibile alle informazioni che contraddicono la convergenza stabilita. La robustezza e l'adeguatezza epistemica possono quindi disaccoppiarsi in modo drammatico: un sistema può produrre output stabilissimi che sono sistematicamente falsi.

Il fenomeno del groupthink, analizzato da Janis (1972) nel contesto delle crisi decisionali politiche americane, è un caso di bolla epistemica prodotta da meccanismi psicologici e organizzativi specifici: pressioni alla conformità all'interno di gruppi coesi, illusione di unanimità, razionalizzazione collettiva delle decisioni già prese, stereotipizzazione dei dissenzienti come deboli o incompetenti. I casi analizzati da Janis — la pianificazione dell'invasione della Baia dei Porci, la gestione della crisi coreana, le decisioni sull'escalation in Vietnam — mostrano come sistemi distribuiti di alta qualità individuale (i decision-makers coinvolti erano spesso esperti di grande competenza) possano produrre atti linguistici collettivi — piani, valutazioni, dichiarazioni pubbliche — epistemicamente degradati per ragioni strutturali legate alla dinamica del gruppo.

La diagnosi del groupthink in termini di struttura della funzione di convergenza è precisa: il processo di convergenza assegna pesi pragmatici molto elevati alla coesione del gruppo e molto bassi alle voci dissenzienti, producendo un output che rispecchia le preferenze del nucleo del gruppo indipendentemente dall'adeguatezza epistemica complessiva.

VI. La verità di sistema: un concetto epistemologico nuovo

Le considerazioni sviluppate nelle sezioni precedenti suggeriscono la necessità di introdurre un concetto epistemologico che la tradizione non ha articolato: la verità di sistema, distinta sia dalla verità individuale (ciò che il singolo agente crede giustificatamente) sia dalla verità procedurale (ciò che le procedure del sistema certificano come vero).

La verità di sistema di una proposizione p rispetto a un sistema distribuito $\mathcal{S}$ può essere approssimativamente definita come: p è vera-di-sistema per $\mathcal{S}$ se p è epistemicamente giustificata dalla totalità delle informazioni disponibili negli agenti di $\mathcal{S}$, integrate in modo epistemicamente appropriato. L'ultima clausola è cruciale: non si tratta semplicemente dell'unione delle credenze degli agenti, ma di una loro integrazione che rispetti certi criteri di adeguatezza epistemica — che dia il peso giusto alle informazioni rilevanti, che tenga conto della loro qualità, che non sia distorta da incentivi non epistemici.

La verità di sistema è diversa dalla verità procedurale (quella prodotta dalla funzione di convergenza effettiva del sistema) perché il sistema può convergere su affermazioni che non sono epistemicamente giustificate dalla totalità delle sue informazioni. Ed è diversa dalla semplice unione delle credenze individuali perché quest'ultima non è generalmente coerente — gli agenti individuali possono avere credenze incompatibili — e non tiene conto dei pesi differenziali delle informazioni.

Implicazioni per la responsabilità epistemica. Il concetto di verità di sistema ha conseguenze dirette per la teoria della responsabilità degli atti linguistici distribuiti. Un'asserzione istituzionale è epistemicamente responsabile quando si avvicina alla verità di sistema del sistema che la produce. La distanza tra verità di sistema e output della convergenza effettiva è una misura del deficit epistemico dell'atto linguistico — della

misura in cui il processo di convergenza ha fallito nel mobilizzare adeguatamente le risorse cognitive del sistema.

Questa misura di deficit epistemico è normativa ma non moralizzante: un deficit elevato può essere il risultato di un processo di convergenza mal progettato (responsabilità strutturale), di comportamenti opportunistici di certi agenti (responsabilità individuale), o di limiti cognitivi del sistema che nessuna progettazione potrebbe eliminare completamente (limite intrinseco). Distinguere tra queste cause è essenziale per attribuire responsabilità in modo appropriato.

VII. Epistemologia dell'intelligenza artificiale distribuita

Il caso dei sistemi di intelligenza artificiale generativa pone il problema dell'epistemologia del testimone distribuito nella sua forma più radicale e più urgente. Un modello linguistico di grandi dimensioni produce atti assertivi — risponde a domande, fornisce informazioni, esprime valutazioni — senza avere credenze nel senso psicologico ordinario. Eppure i suoi output hanno una struttura che può essere analizzata in termini epistemici: sono giustificati o ingiustificati, accurati o inaccurati, coerenti o contraddittori.

La domanda epistemologica fondamentale riguarda la natura della "conoscenza" di un modello linguistico. La risposta più comune è quella distributivista: il modello non "sa" nulla in senso proprio, ma le distribuzioni di probabilità sui token che produce riflettono le strutture statistiche del corpus su cui è stato addestrato. Questa risposta è tecnicamente accurata ma epistemologicamente insoddisfacente: lascia senza risposta la domanda su come debbano essere valutate le asserzioni del modello e su come si debba reagire alle sue "credenze" false o ingiustificate.

Un'alternativa più fruttuosa, coerente con il framework della pragmatica distribuita, è quella di concepire il modello come un sistema di testimonianza distribuita che aggrega le "credenze" di milioni di testi umani in una struttura probabilistica. In questa prospettiva, la "conoscenza" del modello è una forma di conoscenza testimoniale di secondo ordine: riflette non direttamente la realtà ma la distribuzione delle credenze espresse nei testi che ha processato. Questa caratterizzazione ha importanti implicazioni:

Prima implicazione: La qualità epistemica degli output di un modello dipende dalla qualità epistemica del corpus di addestramento. Un corpus che sovrarappresenta certe voci, certe culture, certi periodi storici produce un modello che aggrega testimonianza distribuita in modo epistemicamente distorto. Le distorsioni dei modelli linguistici rispetto a certi gruppi demografici e culturali — documentate da una crescente letteratura — possono essere concettualizzate come forme di ingiustizia testimoniale distribuita incorporate nella struttura della funzione di convergenza del modello.

Seconda implicazione: La relazione tra output del modello e verità di sistema è mediata da due passaggi: dalla realtà alla distribuzione delle credenze espresse nel corpus, e dalla distribuzione delle credenze nel corpus all'output del modello. Ciascun passaggio introduce distorsioni che si sommano. La prima distorsione riflette la qualità epistemica del corpus come sistema di testimonianza distribuita; la seconda riflette le proprietà dell'algoritmo di addestramento come funzione di aggregazione delle credenze.

Terza implicazione: Le pratiche di "hallucination" dei modelli linguistici — la produzione di affermazioni false con tono assertivo sicuro — non sono semplici errori computazionali ma manifestazioni di una struttura epistemica specifica: il modello produce output che massimizzano la coerenza locale (ogni parola è plausibile dato il contesto) senza meccanismi che verifichino la corrispondenza con la realtà esterna al corpus. È la versione computazionale del groupthink: una convergenza robusta su un output internamente coerente che può essere sistematicamente falso.

VIII. Verso un'epistemologia normativa del distribuito

Le considerazioni sviluppate in questa parte convergono verso l'esigenza di un'epistemologia normativa specificamente progettata per sistemi distribuiti — una teoria che dica non solo come la conoscenza si produce nei sistemi collettivi, ma come dovrebbe prodursi per essere epistemicamente adeguata e politicamente legittima.

I principi di tale epistemologia normativa possono essere provvisoriamente enunciati come segue.

Principio di mobilizzazione. Una struttura di convergenza è epistemicamente adeguata nella misura in cui mobilizza la totalità delle informazioni rilevanti disponibili nel sistema, assegnando a ciascuna un peso proporzionale alla sua qualità epistemica piuttosto che alla posizione gerarchica o allo status del suo portatore.

Principio di diversità epistemica. L'adeguatezza epistemica di un sistema distribuito cresce con la diversità delle prospettive epistemiche dei suoi agenti. Sistemi composti da agenti con formazioni, esperienze e quadri concettuali eterogenei tendono a produrre convergenze più robuste e meno soggette a cascate informative rispetto a sistemi omogenei, perché la diversità introduce resistenza strutturale alla conformità.

Principio di tracciabilità. Una struttura di convergenza è epistemicamente responsabile nella misura in cui il percorso che ha portato dall'informazione distribuita all'output finale è ricostruibile e analizzabile. La tracciabilità della convergenza non è solo una condizione di trasparenza politica ma una condizione di responsabilità epistemica: consente di identificare dove si sono prodotte le distorsioni e chi ne è responsabile.

Principio di revisione. Un sistema epistemicamente adeguato deve avere meccanismi strutturati per la revisione degli atti assertivi alla luce di nuove informazioni. La rigidità della convergenza — la difficoltà di modificare l'output una volta raggiunto l'accordo — è una virtù procedurale (garantisce la certezza del diritto, la stabilità degli accordi) ma un vizio epistemico (impedisce il recepimento di nuove evidenze). Il design delle strutture di convergenza deve trovare un equilibrio tra queste due esigenze, che in molti sistemi istituzionali si rispecchia nella distinzione tra atti revocabili e irrevocabili, tra pronunciamenti provvisori e definitivi.

Principio di protezione del dissenso. Le voci dissenzienti all'interno di un sistema distribuito hanno un valore epistemico che spesso supera il loro peso pragmatico nella funzione di convergenza: sono i sensori più sensibili alle discrepanze tra la convergenza stabilita e la verità di sistema. Una struttura di convergenza epistemicamente adeguata deve proteggere il dissenso non solo come diritto individuale ma come risorsa epistemica collettiva — deve costruire meccanismi che consentano alle voci dissenzienti di incidere sull'output anche quando sono in minoranza.

Questi principi non sono una teoria completa: sono i pilastri di un programma di ricerca che richiede sviluppo teorico e verifica empirica. Ma hanno già, nella forma abbozzata, conseguenze pratiche rilevanti per il design delle strutture istituzionali, la valutazione delle pratiche comunicative delle organizzazioni, e — forse in modo più urgente — la progettazione e la regolazione dei sistemi di intelligenza artificiale che producono atti linguistici su scala globale.

Etica delle strutture di convergenza

I. Perché le strutture di convergenza hanno rilevanza etica

L'etica del discorso — dalla tradizione kantiana alle sue elaborazioni contemporanee in Habermas, Apel e Scanlon — ha sempre trattato la dimensione morale della comunicazione come una questione di principi che governano le relazioni tra soggetti: principi di onestà, di rispetto, di non manipolazione, di reciprocità. In questa prospettiva, l'azione comunicativa moralmente corretta è quella che rispetta l'autonomia e la razionalità dell'interlocutore, che non distorce il processo comunicativo attraverso la menzogna o la manipolazione, che è orientata al raggiungimento del consenso argomentato piuttosto che al perseguimento di interessi particolari.

Questa tradizione ha prodotto risultati importanti per l'analisi dell'etica della comunicazione interpersonale. Ma rimane, nella sua impostazione fondamentale, vincolata al modello del soggetto unitario: i principi etici che elabora si applicano a individui capaci di intenzioni morali, di deliberazione e di responsabilità. Quando il soggetto comunicativo è distribuito — quando è un'organizzazione, un'istituzione, un sistema computazionale — la questione etica si sposta dal livello dei principi individuali al livello delle strutture: non è sufficiente che i singoli agenti si comportino eticamente; occorre che le strutture attraverso cui i loro contributi vengono integrati siano esse stesse eticamente adeguate.

Questa è la tesi centrale dell'etica delle strutture di convergenza: le strutture istituzionali, organizzative e computazionali attraverso cui i sistemi distribuiti producono atti linguistici hanno una rilevanza etica autonoma che non è riducibile alla somma delle valutazioni etiche dei comportamenti individuali. Una struttura di convergenza può essere

eticamente problematica anche quando tutti i suoi agenti si comportano individualmente in modo eticamente corretto; e può produrre output eticamente positivi anche in presenza di alcuni agenti con intenzioni problematiche, se i meccanismi di convergenza sono progettati per mitigare queste distorsioni.

II. Il male strutturale nella pragmatica distribuita

Il concetto di male strutturale — l'idea che certi danni siano prodotti non da intenzioni malvagie individuali ma da strutture che generano sistematicamente certi effetti indipendentemente dalle intenzioni dei partecipanti — ha radici nelle teorie critiche della società, da Marx alla tradizione femminista, dai Critical Race Studies alla filosofia analitica della giustizia sociale. Iris Marion Young ha articolato questa nozione in modo particolarmente rigoroso nel suo saggio sulla responsabilità sociale per l'ingiustizia strutturale, distinguendo tra la responsabilità individuale per azioni intenzionali e la responsabilità politica per le strutture in cui si partecipa.

La pragmatica distribuita consente di applicare questa nozione con precisione al dominio degli atti linguistici. Il male strutturale comunicativo è la produzione sistematica di atti linguistici che producono danni — epistemici, pratici, morali — attraverso strutture di convergenza che nessun singolo agente ha inteso progettare per produrre quel tipo di danno, e che possono continuare a produrlo anche dopo che i singoli agenti ne sono diventati consapevoli, perché il cambiamento richiederebbe una trasformazione strutturale che nessun singolo agente ha il potere o l'incentivo di realizzare unilateralmente.

I casi esaminati nella quinta parte offrono esempi paradigmatici di male strutturale comunicativo. La produzione sistematica di ratings gonfiati da parte delle agenzie durante la bolla finanziaria non richiedeva che nessun analista avesse l'intenzione di ingannare gli investitori: era sufficiente che la struttura degli incentivi e dei processi di convergenza orientasse l'output verso certe caratteristiche sistematicamente distorte. Analogamente, la produzione di comunicazioni istituzionali che normalizzano discriminazioni di genere o razza non richiede che nessun comunicatore abbia intenzioni discriminatorie: è sufficiente che le strutture di convergenza che producono questi atti linguistici siano composte da agenti che condividono certe assunzioni implicite, che certi

tipi di esperienza siano sistematicamente assenti dal processo di convergenza, e che le procedure di validazione interna non includano meccanismi di verifica critica delle assunzioni implicite.

La responsabilità per il male strutturale comunicativo non può essere modellata secondo lo schema della responsabilità individuale classica (chi ha inteso produrre il danno? chi aveva il potere di evitarlo?), perché le risposte a queste domande non individuano un agente singolo. Deve invece essere modellata secondo lo schema della responsabilità politica di Young: chiunque partecipi a strutture che producono sistematicamente danni ha una responsabilità di lavorare per la loro trasformazione, proporzionale alla propria posizione nel sistema e al proprio potere di incidere sulla struttura.

III. Principi etici per il design delle strutture di convergenza

Il riconoscimento della rilevanza etica delle strutture di convergenza suggerisce la necessità di principi etici che governino non solo il comportamento degli agenti individuali nei processi comunicativi ma il design delle strutture stesse. Propongo sei principi, organizzati in due gruppi: principi proceduali (che riguardano la struttura del processo) e principi sostanziali (che riguardano la qualità epistemica e l'impatto dell'output).

Principi procedurali:

Principio di inclusività. Le strutture di convergenza che producono atti linguistici con forza pubblica significativa devono includere, o dare accesso significativo alle preferenze di, tutti gli agenti che sono sostanzialmente interessati dall'output. Questo principio è l'estensione al dominio comunicativo del principio democratico di inclusività: le decisioni che riguardano una comunità devono essere prodotte da processi in cui quella comunità ha voce.

L'applicazione di questo principio alle strutture di convergenza istituzionali reali è più complessa di quanto possa sembrare. "Dare accesso alle preferenze" non significa necessariamente includere tutti gli interessati nel processo deliberativo formale: può significare costruire meccanismi di consultazione, di rappresentanza, di feedback che consentano alle preferenze degli interessati di influenzare la convergenza

anche senza partecipazione diretta. Ma la complessità dell'applicazione non diminuisce la forza normativa del principio: strutture di convergenza che sistematicamente escludono le voci degli interessati producono output che, indipendentemente dalla loro validità formale, mancano di legittimità comunicativa.

Principio di proporzionalità del peso. Il peso pragmatico di ciascun agente nella funzione di convergenza deve essere proporzionale alla sua competenza epistemica rilevante per il tipo di atto linguistico prodotto, e non deve essere determinato da fattori eticamente irrilevanti come lo status gerarchico, l'identità di gruppo, o la capacità di esercitare pressioni di potere.

Questo principio non è egualitarista in senso semplice: non sostiene che tutte le voci debbano avere lo stesso peso. Sostiene che il peso debba essere determinato da criteri epistemicamente giustificati — expertise, accesso all'informazione rilevante, qualità dell'argomentazione — piuttosto che da fattori che non hanno rilevanza epistemica. In pratica, questo principio richiede che le strutture di convergenza rendano espliciti i criteri con cui assegnano pesi pragmatici agli agenti e che questi criteri siano giustificabili pubblicamente.

Principio di trasparenza della struttura. Le strutture di convergenza che producono atti linguistici con forza pubblica devono essere conoscibili dai destinatari degli atti. I destinatari di un'asserzione istituzionale hanno il diritto di sapere — almeno nelle linee essenziali — come è stata prodotta: chi ha partecipato al processo di convergenza, quali procedure sono state seguite, quali voci sono state incluse o escluse e per quali ragioni.

Questo principio ha fondamento in ragioni di autonomia (i destinatari degli atti linguistici hanno diritto alle informazioni necessarie per valutare criticamente ciò che ricevono) e in ragioni epistemiche (la conoscenza della struttura della convergenza è necessaria per una corretta interpretazione degli atti prodotti). È diverso dal principio di tracciabilità epistemica introdotto nella settima parte: la trasparenza della struttura è un requisito etico che si applica anche quando la tracciabilità non è necessaria per ragioni epistemiche.

Principi sostanziali:

Principio di adeguatezza epistemica. Le strutture di convergenza devono essere progettate per massimizzare la qualità epistemica degli output, mobilizzando le informazioni rilevanti disponibili nel sistema e minimizzando le distorsioni prodotte da incentivi non epistemici. Una struttura di convergenza che sistematicamente produce output epistemicamente degradati è eticamente problematica, anche quando è proceduralmente valida, perché produce atti linguistici che danneggiano i destinatari attraverso la trasmissione di informazioni false o fuorvianti.

Principio di non manipolazione strutturale. Le strutture di convergenza non devono essere progettate per produrre output predeterminati indipendentemente dalle informazioni e dai contributi degli agenti partecipanti. Una struttura di convergenza che simula la deliberazione collettiva pur essendo orientata a produrre l'output voluto dal progettista è una forma di manipolazione istituzionale che viola l'autonomia degli agenti partecipanti e la fiducia dei destinatari.

Questo principio è particolarmente rilevante per la progettazione dei sistemi di intelligenza artificiale: un sistema di AI progettato per produrre certi tipi di output attraverso tecniche di allineamento che orientano la distribuzione di probabilità verso risposte predeterminate, senza che l'utente sia consapevole di questa struttura, è una forma di manipolazione strutturale. La trasparenza sull'allineamento — su come il sistema è stato progettato per comportarsi e perché — è un requisito etico che discende direttamente da questo principio.

Principio di proporzionalità dell'impatto. I requisiti etici per le strutture di convergenza sono proporzionali all'impatto degli atti linguistici che producono. Strutture che producono atti linguistici con forza normativa ampia — sentenze, leggi, politiche pubbliche, comunicazioni di massa — devono soddisfare standard etici più stringenti di strutture che producono atti linguistici con forza limitata. Questo principio di proporzionalità riconosce che non tutti i sistemi distribuiti hanno le risorse o la necessità di soddisfare gli standard massimi, e che i costi del design etico devono essere calibrati in relazione ai benefici attesi.

IV. Giustizia comunicativa come categoria normativa

Il framework dei principi etici sviluppato nella sezione precedente può essere integrato in una nozione più ampia di giustizia comunicativa:

l'insieme delle condizioni strutturali che devono essere soddisfatte affinché i sistemi distribuiti producano atti linguistici in modo giusto. La giustizia comunicativa è distinta dalla giustizia distributiva (la giusta distribuzione dei beni) e dalla giustizia procedurale (la correttezza delle procedure) sebbene abbia relazioni con entrambe.

Il concetto ha precedenti nella filosofia politica del linguaggio. La già citata nozione di ingiustizia epistemica di Fricker identifica alcune condizioni di giustizia nelle relazioni comunicative interpersonali. La teoria dell'azione comunicativa di Habermas articola condizioni ideali di giustizia per i processi deliberativi. La critica del discorso di Foucault identifica relazioni di potere incorporati nelle strutture discorsive. Ma nessuna di queste tradizioni ha articolato sistematicamente la giustizia comunicativa come proprietà delle strutture di convergenza.

Una prima articolazione può essere proposta attraverso tre dimensioni:

Giustizia nell'accesso. Un sistema distributivo produce atti linguistici in modo giusto rispetto alla dimensione dell'accesso quando tutti gli agenti legittimamente interessati hanno la possibilità di contribuire al processo di convergenza con peso appropriato alla loro competenza e ai loro interessi. Le violazioni della giustizia nell'accesso includono: l'esclusione deliberata di certi agenti dal processo; l'assegnazione di pesi pragmatici sproporzionatamente bassi a certi tipi di voci; la strutturazione delle procedure in modo da rendere difficile la partecipazione per certi agenti (barriere linguistiche, temporali, informative); la mancanza di meccanismi che garantiscano la rappresentanza di interessi diffusi di fronte a interessi concentrati.

Giustizia nella distribuzione degli effetti. Un sistema distribuito produce atti linguistici in modo giusto rispetto alla dimensione degli effetti quando i benefici e i costi degli atti prodotti sono distribuiti in modo equo tra tutti gli interessati, e quando le strutture di convergenza non sono progettate — esplicitamente o implicitamente — per concentrare i benefici su certi agenti a spese di altri. Le violazioni della giustizia nella distribuzione degli effetti includono: la produzione sistematica di atti linguistici che normalizzano o rafforzano ineguaglianze preesistenti; la strutturazione delle procedure in modo da sistematicamente favorire certi interessi; il disegno di strutture di convergenza che rendono invisibili certi tipi di effetti negativi o che li scaricano su soggetti privi di voce nel processo.

Giustizia nel riconoscimento. Un sistema distribuito produce atti linguistici in modo giusto rispetto alla dimensione del riconoscimento quando le strutture di convergenza riconoscono la piena dignity e agency epistemica di tutti gli agenti partecipanti, senza subordinare sistematicamente certe voci o certe categorie di esperienza. Le violazioni della giustizia nel riconoscimento includono: la sistematica squalificazione di certe forme di conoscenza come non pertinenti o non affidabili per ragioni legate all'identità del portatore piuttosto che alla qualità epistemica; la strutturazione delle procedure in modo da rendere difficile la comunicazione di certe categorie di esperienza; il design di sistemi artificiali che riproducono e amplificano distorsioni di riconoscimento presenti nel corpus di addestramento.

V. Il problema della responsabilità collettiva

Una delle questioni etiche più complesse che emergono dalla pragmatica distribuita riguarda la responsabilità collettiva: in che senso un sistema distribuito nel suo complesso può essere ritenuto responsabile degli atti linguistici che produce? Questa domanda ha implicazioni sia teoriche — riguarda i fondamenti della responsabilità collettiva — sia pratiche, perché da essa dipende la possibilità di attribuire responsabilità giuridica agli enti collettivi per danni prodotti attraverso atti linguistici.

La tradizione filosofica sulla responsabilità collettiva è divisa tra posizioni reductiviste (che sostengono che la responsabilità degli enti è sempre riducibile alla responsabilità degli individui che li compongono) e posizioni emergentiste (che sostengono che gli enti collettivi possono avere responsabilità propria non riducibile a quella degli individui). Il dibattito ha implicazioni pratiche rilevanti: se la responsabilità collettiva è riduzionista, sanzionare l'ente è solo un modo indiretto di sanzionare gli individui; se è emergente, sanzionare l'ente ha una giustificazione normativa propria.

La pragmatica dell'intenzione distribuita offre una prospettiva originale su questo dibattito. La responsabilità dell'ente per i suoi atti linguistici non è né puramente riduzionista né puramente emergente: è strutturalmente dipendente, nel senso che dipende dalla struttura della funzione di convergenza che ha prodotto l'atto. Un ente è responsabile nella misura in cui:

1. la struttura di convergenza che ha prodotto l'atto era sotto il suo controllo normativo (poteva essere progettata diversamente);
2. la struttura ha effettivamente prodotto l'atto in modo causalmente rilevante;
3. l'ente ha avuto la possibilità di rivedere l'atto o di prevenirne gli effetti dannosi.

Questa definizione di responsabilità strutturale dell'ente è più precisa della responsabilità per "colpa di organizzazione" del diritto penale degli enti (che richiede l'identificazione di un difetto organizzativo specifico) e più ricca della semplice responsabilità oggettiva (che non richiede nessuna relazione tra la struttura interna e il danno prodotto).

Il problema della continuità dell'ente. Una complicazione specifica della responsabilità collettiva riguarda la continuità nel tempo: un ente che ha prodotto atti linguistici dannosi in passato rimane responsabile dopo che gli agenti che partecipavano al processo di convergenza sono stati sostituiti? La pragmatica distribuita suggerisce che la risposta dipenda dalla continuità delle strutture di convergenza: se le strutture che hanno prodotto gli atti dannosi sono ancora in funzione, la responsabilità dell'ente persiste anche al cambiare degli individui; se le strutture sono state trasformate in modo da non poter più produrre quel tipo di output, la responsabilità dell'ente si attenua, anche se rimangono obblighi di riparazione per i danni passati.

VI. Etica della progettazione dei sistemi AI come etica della convergenza

Nessuna applicazione dell'etica delle strutture di convergenza è più urgente — e nessuna è più controversa — di quella che riguarda la progettazione dei sistemi di intelligenza artificiale generativa. Questi sistemi stanno diventando agenti linguistici di forza crescente nella sfera pubblica: forniscono informazioni, formulano consigli, modellano opinioni, producono contenuti che circolano come discorso umano. L'etica della loro progettazione è, dal punto di vista della pragmatica distribuita, un caso speciale dell'etica del design delle strutture di convergenza.

I principi enunciati nella terza sezione si applicano con conseguenze specifiche.

Il *principio di inclusività* richiede che il corpus di addestramento e le procedure di fine-tuning includano rappresentazioni adeguate della diversità delle voci, delle culture, delle prospettive epistemiche che costituiscono l'umanità. Un sistema addestrato prevalentemente su testi in lingua inglese, prodotti da autori di certi profili demografici e culturali, su piattaforme con certi pattern di selezione, è una struttura di convergenza che viola sistematicamente il principio di inclusività — non per intenzione dei progettisti, ma per le scelte strutturali incorporate nel design del processo di addestramento.

Il *principio di proporzionalità del peso* richiede che le tecniche di allineamento che modellano il comportamento del sistema non assegnino pesi sproporzionati alle preferenze di certi tipi di annotatori (tipicamente, pool di annotatori relativamente omogenei per demografica e cultura) rispetto alla varietà delle prospettive umane. Le distorsioni sistematiche nei sistemi di AI — che tendono a riprodurre e amplificare le preferenze dei gruppi sovrarappresentati nel processo di allineamento — sono violazioni strutturali di questo principio.

Il *principio di trasparenza della struttura* richiede che gli utenti dei sistemi AI abbiano accesso a informazioni significative su come il sistema è stato progettato: quali corpus, quali tecniche di allineamento, quali valori e preferenze sono stati incorporati nel sistema e perché. Questa trasparenza non è solo una condizione di autonomia degli utenti ma una condizione epistemica: senza conoscenza della struttura di convergenza che ha prodotto un output, i destinatari non hanno gli strumenti per valutarne criticamente la qualità e l'affidabilità.

Il *principio di non manipolazione strutturale* ha implicazioni particolarmente stringenti per i sistemi AI. Un sistema progettato per produrre certi tipi di output attraverso tecniche di allineamento che non sono trasparenti all'utente è una macchina di manipolazione strutturale: produce atti linguistici che sembrano emerge naturalmente da un processo di elaborazione dell'informazione, ma che riflettono scelte normative incorporate nel design del sistema. La distinzione tra allineamento legittimo (che incide il sistema a rispettare valori etici fondamentali) e manipolazione strutturale (che orienta il sistema verso preferenze non dichiarate dei progettisti o dei committenti) è sottile ma cruciale, e la sua articolazione è uno dei compiti più urgenti dell'etica applicata dell'AI.

VII. Il dovere di design etico: chi è responsabile delle strutture?

Un problema che l'etica delle strutture di convergenza deve affrontare riguarda la distribuzione della responsabilità per il design: chi ha il dovere di assicurare che le strutture di convergenza siano eticamente adeguate? La risposta non può essere "tutti" in modo generico, perché questo diluisce la responsabilità al punto da renderla praticamente inesigibile. Deve essere distribuita in modo proporzionale alla posizione e al potere.

I progettisti delle strutture di convergenza — i legislatori che scrivono le procedure istituzionali, i manager che definiscono i processi organizzativi, gli ingegneri che progettano i sistemi computazionali — hanno il dovere primario di design etico. Hanno la conoscenza tecnica necessaria per comprendere come le strutture che progettano influenzeranno gli output, e hanno il potere di scegliere tra design alternativi. La responsabilità per design della convergenza — introdotta nella quarta parte come terzo livello del framework di responsabilità — è precisamente la responsabilità di questi agenti.

I partecipanti ai processi di convergenza — gli agenti che contribuiscono alla produzione degli atti linguistici all'interno di strutture esistenti — hanno un dovere derivato di segnalare e resistere alle distorsioni strutturali che identificano nel processo. Non hanno il potere di cambiare unilateralmente le strutture, ma hanno la responsabilità di non partecipare silenziosamente alla produzione di output eticamente problematici e di segnalare le distorsioni attraverso i canali disponibili.

I destinatari degli atti linguistici distribuiti hanno un dovere di scrutinio critico delle strutture di convergenza che producono i discorsi istituzionali a cui sono esposti. Questo dovere non è semplicemente individuale: richiede strutture sociali e politiche che rendano possibile lo scrutinio — libertà di stampa, accesso all'informazione, istituzioni di accountability indipendenti. In questo senso, il dovere di scrutinio critico delle strutture di convergenza è anche un dovere collettivo di mantenere le condizioni istituzionali che rendono possibile tale scrutinio.

Le istituzioni regolatorie hanno il dovere di stabilire standard minimi per le strutture di convergenza che producono atti linguistici ad alto impatto pubblico e di farli rispettare. Questo dovere si estende — e questo è uno dei contributi più originali della prospettiva qui proposta — ben oltre i tradizionali ambiti della regolazione del discorso pubblico

(regolazione della stampa, della pubblicità, del discorso politico) per abbracciare la regolazione delle strutture organizzative e computazionali attraverso cui questo discorso viene prodotto.

VIII. Il limite etico dell'efficienza comunicativa

Un'ultima questione etica merita attenzione: il rapporto tra efficienza comunicativa e giustizia comunicativa. Le strutture di convergenza eticamente adeguate sono, tipicamente, più costose di quelle eticamente inadeguate: includere più voci, assegnare pesi equi, garantire trasparenza, proteggere il dissenso — tutto ciò richiede risorse, tempo e complessità procedurale. L'efficienza comunicativa — la capacità di produrre rapidamente output con forza illocutoria determinata — tende a spingere verso strutture di convergenza gerarchiche, opache e semplificatrici.

La tensione tra efficienza e giustizia comunicativa non è risolvibile in astratto: è una delle tensioni fondamentali del design istituzionale, analoga alla tensione tra efficienza e giustizia nella distribuzione delle risorse economiche. Ma il riconoscimento di questa tensione ha un valore teorico importante: rende visibile il fatto che le strutture di convergenza efficienti hanno costi etici reali che tendono a essere invisibili nella contabilità istituzionale ordinaria, e che la scelta di strutture efficienti a scapito di strutture giuste non è una scelta tecnica neutra ma una scelta normativa che richiede giustificazione.

In particolare, il costo dell'efficienza tende a essere sistematicamente distribuito in modo ineguale: i benefici dell'efficienza comunicativa (output rapidi e certi) vanno tipicamente a chi ha già posizioni di potere nel sistema, mentre i costi (esclusione, sottorappresentazione, mancanza di trasparenza) ricadono tipicamente su chi è già ai margini del processo di convergenza. Questa distribuzione ineguale dei costi dell'efficienza è una forma di ingiustizia strutturale comunicativa che l'etica delle strutture di convergenza ha il compito di rendere visibile e di resistere.

La conclusione pragmatica di questa parte non è utopica: non sostiene che ogni struttura di convergenza debba massimizzare simultaneamente tutti i principi etici, il che sarebbe sia impossibile (per il teorema di Arrow) sia impraticabile (per vincoli di risorse). Sostiene che i trade-off etici incorporati nel design delle strutture di convergenza siano resi espliciti, giustificati pubblicamente, e sottoposti a scrutinio critico da

parte di coloro che ne subiscono i costi. Questa è la condizione minima — e irrinunciabile — di un'etica delle strutture di convergenza che sia qualcosa di più di una razionalizzazione di poteri già costituiti.

Una semiotica del distribuito

I. Il problema semiotico del soggetto assente

La pragmatica e la semiotica hanno avuto storicamente rapporti di tensione produttiva. La semiotica — nella tradizione saussuriana come in quella peirciana — si è occupata dei meccanismi attraverso cui i segni producono significato indipendentemente dal soggetto che li usa; la pragmatica — nella tradizione che si è analizzata nelle parti precedenti — si è occupata di come i soggetti usano i segni per compiere azioni. La differenza di enfasi non è irrilevante: la semiotica tende a privilegiare la struttura del sistema significante rispetto all'agente che lo usa; la pragmatica tende a privilegiare l'agente rispetto alla struttura. La pragmatica dell'intenzione distribuita, avendo rimosso il soggetto unitario come fondamento, si trova in una posizione teorica in cui la distinzione classica tra semiotica e pragmatica deve essere rinegoziata.

La domanda semiotica fondamentale che emerge dalla rimozione del PSU è questa: come funziona la referenza in assenza di un soggetto unitario? Cosa significa, per un sistema distribuito, usare un segno per riferirsi a qualcosa? E come si determina il significato di un atto linguistico prodotto da un sistema in cui non esiste un centro interpretativo che abbia accesso all'intera rete di significati mobilitata dall'atto?

Queste domande non hanno risposte ovvie, e alcune tradizioni semiotiche offrono risorse più adeguate di altre per affrontarle. In questa parte si esamineranno tre tradizioni — la semiotica peirciana, la semiotica del testo di matrice europea, e la sociosemiotica — per estrarne gli strumenti più rilevanti per una semiotica del distribuito, e si articoleranno alcune proposte teoriche originali che integrano questi strumenti con il framework della funzione di convergenza.

II. Peirce rivisitato: la semiosi come processo distribuito

Charles Sanders Peirce ha costruito la semiotica più sofisticata e più feconda della tradizione occidentale, e ha avuto la lungimiranza di costruirla senza ancorare il processo semiotico a un soggetto psicologico unitario. La semiosi peirciana — il processo attraverso cui un segno rinvia a un oggetto attraverso un interpretante — è già, nella concezione originale, un processo potenzialmente distribuito: l'interpretante non è necessariamente una mente individuale ma qualunque "effetto del segno", che può realizzarsi in una mente individuale, in una abitudine comportamentale, in un altro segno.

La struttura triadica del segno peirciano — Representamen (il veicolo segnico), Oggetto (ciò a cui il segno si riferisce), Interpretante (l'effetto del segno) — è crucialmente diversa dalla struttura diadica saussuriana (significante/significato) proprio perché include la dimensione dell'effetto: il segno non ha significato in virtù di una relazione interna tra i suoi componenti ma in virtù della sua capacità di produrre effetti interpretativi. Questa struttura è già un'apertura verso una semiotica del distribuito: il significato di un segno è distribuito attraverso tutti gli interpretanti che produce, e non è mai completamente localizzabile in nessuno di essi.

L'interpretante finale come convergenza. Peirce distingue tra interpretanti immediati (il significato che il segno ha in virtù della sua struttura interna), dinamici (gli effetti reali che il segno produce in interpreti specifici) e finali (il significato che il segno avrebbe se l'interpretazione fosse condotta fino in fondo, all'ideale limite della comunità di indagine). L'interpretante finale è una nozione regolativa: non è il significato che un soggetto o anche una comunità attuale attribuisce al segno, ma il significato verso cui la comunità di indagine tenderebbe nell'ipotetico limite dell'investigazione completa.

Questa nozione di interpretante finale può essere riletta, nella prospettiva della pragmatica distribuita, come una forma di funzione di convergenza ideale: il significato verso cui converge il processo distributivo di interpretazione quando tutti i vincoli epistemici e comunicativi sono pienamente soddisfatti. La convergenza reale — quella prodotta dalle strutture di convergenza effettive dei sistemi distribuiti — è sempre un'approssimazione imperfetta di questa convergenza ideale, tanto più imperfetta quanto più le strutture di convergenza si discostano dai

principi di adeguatezza epistemica e inclusività articolati nella parte precedente.

L'indicalità e il problema della referenza distribuita. Una delle tassonomie peirciane più rilevanti per la pragmatica distribuita è quella che distingue i segni in icone (segni che riferiscono per somiglianza con l'oggetto), indici (segni che riferiscono per connessione causale o spaziotemporale con l'oggetto) e simboli (segni che riferiscono per convenzione). Gli atti linguistici in senso stretto appartengono prevalentemente alla classe dei simboli: le parole e le locuzioni convenzionali rinviano ai loro referenti in virtù di convenzioni linguistiche condivise dalla comunità dei parlanti.

Ma molti atti linguistici istituzionali hanno anche una forte componente indicale: rinviano a referenti specifici non solo in virtù del loro contenuto semantico ma in virtù della loro posizione nel contesto istituzionale. Una sentenza che condanna "l'imputato nel procedimento X" ha forza indicale: rinvia a una persona specifica non solo attraverso la descrizione ma attraverso la posizione che il testo occupa nel procedimento giudiziario. Questo significa che la referenza di un atto linguistico istituzionale non è determinata solo dalla sua struttura semantica ma anche dalla sua posizione nel sistema istituzionale — dalla rete di pratiche, procedure e relazioni che costituisce il contesto in cui l'atto è prodotto.

La referenza distribuita è la situazione in cui la componente indicale della referenza è essa stessa distribuita: non c'è un singolo agente che "intende" il referente, ma il referente è determinato dalla posizione dell'atto nella struttura istituzionale. Una delibera del consiglio di amministrazione che dispone "il trasferimento delle attività alla controllata Y" non ha il suo referente determinato dall'intenzione di nessun singolo consigliere: il referente è determinato dall'intero sistema di relazioni giuridiche, contrattuali e organizzative in cui la delibera è inserita. La semiotica del distribuito deve dare conto di questo tipo di referenza senza ricorrere a un soggetto intenzionale che "punta" al referente.

III. Il testo come convergenza: la semiotica di Lotman rivisitata

La semiotica della cultura elaborata da Jurij Lotman e dalla scuola di Tartu-Mosca offre strumenti concettuali diversi ma complementari per una semiotica del distribuito. Lotman concepisce la cultura come una "semiosfera" — uno spazio semiotico eterogeneo in cui testi diversi, codici diversi, lingue diverse coesistono in relazioni di traduzione reciproca e di tensione produttiva. Il testo, in questa prospettiva, non è il prodotto di un autore individuale ma un nodo nella rete di relazioni semiotiche che costituisce la semiosfera: il suo significato è determinato non solo dal suo contenuto interno ma dalla sua posizione nelle relazioni intertestuali e intersemiotiche della cultura.

Questa concezione del testo come nodo semiotico distribuito è direttamente rilevante per la pragmatica dell'intenzione distribuita. Gli atti linguistici istituzionali — le leggi, le sentenze, i contratti, le dichiarazioni politiche — sono testi nel senso lotmaniano: il loro significato non è determinato dall'intenzione di nessun autore singolo ma dalla loro posizione nella semiosfera istituzionale. Una legge non significa solo ciò che le parole del testo dicono in isolamento: significa ciò che dicono in relazione a tutta la tradizione giuridica, ai precedenti interpretativi, alle pratiche delle corti, ai commentari dottrinali, alle altre norme del sistema.

La dialettica core/periferia nella semiosfera istituzionale. Lotman distingue nel funzionamento della semiosfera un'area "core" — dove i codici e le convenzioni sono stabilizzati e la traduzione tra testi diversi è regolata da metaregole condivise — e un'area "periferia" — dove i codici sono in formazione, i confini sono porosi e la traduzione è problematica. Questa distinzione ha un analogo nella pragmatica distribuita: le strutture di convergenza istituzionali consolidate (costituzioni, codici giuridici, procedure parlamentari) funzionano come core della semiosfera istituzionale, producendo atti linguistici con forza e significato stabili; le strutture emergenti (nuovi regimi internazionali, nuove tecnologie comunicative, nuovi tipi di attori istituzionali) funzionano come periferia, producendo atti linguistici il cui significato è ancora in formazione e la cui forza è incerta.

I sistemi di intelligenza artificiale generativa si trovano tipicamente in questa periferia semiotica: producono atti linguistici che si appropriano delle forme e dei generi della comunicazione umana, ma la cui posizione nella semiosfera istituzionale — e quindi il cui significato e la cui forza — non è ancora determinata. La questione di come attribuire forza illocutoria agli atti linguistici prodotti dai sistemi AI non è solo una

questione giuridica o etica: è una questione semiotica, che riguarda la posizione di questi atti nella semiosfera e i metaregole attraverso cui vengono interpretati.

IV. La sociosemiotica e la grammatica del distribuito

La sociosemiotica — sviluppata da Halliday a partire dalla linguistica sistemico-funzionale e approfondita da Kress, van Leeuwen e altri nel contesto della semiotica multimodale — offre una terza prospettiva rilevante per una semiotica del distribuito. La sociosemiotica studia come i sistemi semiotici funzionino come risorse socialmente condivise per la costruzione del significato, con attenzione particolare al modo in cui queste risorse riflettono e riproducono relazioni di potere sociale.

La nozione hallidayana di funzioni del linguaggio — ideazionale (costruire rappresentazioni del mondo), interpersonale (gestire relazioni sociali), testuale (organizzare il discorso in testi coerenti) — può essere estesa alla pragmatica distribuita in modo che cattura dimensioni della comunicazione istituzionale che le approcci precedenti non colgono pienamente.

La funzione ideazionale della convergenza. Quando un sistema distribuito produce un atto linguistico, non produce solo un messaggio ma costruisce una rappresentazione del mondo: determina quali aspetti della realtà sono degni di menzione, quali categorie concettuali sono appropriate per descriverla, quali relazioni tra entità sono rilevanti. Questa funzione costruttiva della comunicazione istituzionale è particolarmente rilevante perché le rappresentazioni del mondo incorporate negli atti linguistici istituzionali tendono a naturalizzarsi: vengono percepite come descrizioni della realtà piuttosto che come costruzioni di essa.

La critica ideologica del linguaggio — da Marx a Gramsci a van Dijk — ha sempre insistito su questo punto, ma la pragmatica distribuita consente di precisarlo in termini strutturali: le rappresentazioni del mondo incorporate negli atti linguistici istituzionali riflettono le strutture di convergenza che li hanno prodotti. Se il processo di convergenza include agenti con certe assunzioni implicite sul mondo e ne esclude altri, il risultato sarà sistematicamente orientato verso certe rappresentazioni

del mondo a scapito di altre, non per intenzione di nessun agente ma per la struttura della funzione di convergenza.

La funzione interpersonale della convergenza. La funzione interpersonale del linguaggio costruisce posizioni di identità: chi parla e chi ascolta, chi ha autorità e chi deve rispettarla, chi è incluso nella comunità e chi è escluso. Nella pragmatica distribuita, queste posizioni di identità sono costruite non dai singoli parlanti ma dalle strutture di convergenza che producono gli atti linguistici istituzionali.

Un decreto ministeriale non costruisce solo una disposizione normativa: costruisce anche un sistema di posizioni identitarie — il Ministro come soggetto autorizzato, i cittadini come soggetti obbligati, gli esperti come fonti di giustificazione, le minoranze come soggetti da tutelare o da disciplinare. Queste costruzioni identitarie riflettono le strutture di convergenza che hanno prodotto il decreto: chi aveva voce nel processo, quali interessi sono stati rappresentati, quali categorie concettuali erano disponibili nel sistema.

La sociosemiotica fornisce strumenti analitici precisi per analizzare queste costruzioni — la grammatica della transitività per le costruzioni ideazionali, la grammatica della modalità per le costruzioni interpersonali — che possono essere integrati nel framework della pragmatica distribuita per produrre analisi più complete degli atti linguistici istituzionali.

V. Il problema dell'intertestualità distribuita

Un fenomeno semiotico che la pragmatica distribuita deve affrontare con strumenti propri è quello dell'intertestualità: la relazione tra un testo e gli altri testi che presuppone, cita, contraddice o trasforma. L'intertestualità negli atti linguistici istituzionali è massiccia e strutturale: ogni legge presuppone la tradizione giuridica in cui si inserisce; ogni contratto richiama i precedenti contrattuali e le prassi del settore; ogni dichiarazione politica si posiziona rispetto al repertorio dei discorsi politici disponibili.

La tradizione semiotica ha analizzato l'intertestualità come una proprietà dei testi, indipendentemente dall'intenzione degli autori: un testo è intertestualmente connesso ad altri testi indipendentemente dal fatto che

il suo autore ne fosse consapevole. Questa concezione è direttamente applicabile alla pragmatica distribuita, dove l'intenzione dell'autore è in ogni caso distribuita e parziale: l'intertestualità degli atti linguistici istituzionali è determinata dalla posizione che occupano nella semiosfera istituzionale, non dalla consapevolezza dei singoli agenti che vi hanno contribuito.

L'intertestualità come vincolo sulla funzione di convergenza. Un'implicazione importante di questa prospettiva è che la funzione di convergenza non opera in uno spazio vuoto ma in uno spazio strutturato da relazioni intertestuali preesistenti. I partecipanti al processo di convergenza non possono produrre qualsiasi output: sono vincolati dalle relazioni intertestuali del sistema istituzionale in cui operano. Un giudice non può produrre una sentenza che contraddica apertamente la giurisprudenza consolidata senza dovere giustificare la deviazione; un legislatore non può produrre una legge che ignori le norme costituzionali; un'organizzazione non può produrre una dichiarazione pubblica che contraddica apertamente le sue dichiarazioni precedenti senza produrre una forma di incoerenza che richiede spiegazione.

Questi vincoli intertestuali sono una risorsa — forniscono stabilità e prevedibilità agli atti linguistici istituzionali — ma anche una fonte di inerzia: rendono difficile la produzione di atti linguistici radicalmente nuovi, anche quando la novità sarebbe epistemicamente o eticamente giustificata. La comprensione di questa tensione tra vincoli intertestuali e innovazione discorsiva è uno degli aspetti più importanti di una semiotica del distribuito.

VI. Il significato emergente: dalle interazioni locali ai pattern globali

Una delle caratteristiche più affascinanti dei sistemi semiotici distribuiti è la produzione di significato emergente: significati che non erano intesi da nessun partecipante al processo di convergenza ma che emergono dalle interazioni tra i contributi dei diversi agenti. Questo fenomeno — ben noto agli studiosi di discorso istituzionale — è difficilmente spiegabile dalla pragmatica classica, che richiede che il significato di un atto sia intenzionalmente prodotto dal parlante.

Un esempio paradigmatico è quello della legislazione per accumulo di emendamenti. Un testo legislativo che nasce con un obiettivo dichiarato può, attraverso un processo di emendamento in cui ciascun emendamento individuale ha una giustificazione locale che gli autori possono genuinamente difendere, produrre un testo finale il cui significato complessivo è radicalmente diverso dall'intenzione originaria — e che può persino contraddirla. Il significato emergente del testo finale non è inteso da nessuno dei partecipanti al processo di convergenza, ma è una proprietà reale del testo che ha effetti pratici reali.

Dal punto di vista della pragmatica distribuita, il significato emergente è un caso in cui la funzione di convergenza produce un output le cui proprietà semantiche non sono riducibili alle preferenze e alle intenzioni dei partecipanti. Questo non significa che il significato emergente sia imperscrutabile o inanalizzabile: può essere ricostruito attraverso l'analisi della struttura della funzione di convergenza e delle sue trasformazioni nel tempo. Ma non può essere compreso attraverso la semplice attribuzione di intenzioni al sistema nel suo complesso.

Il significato emergente nei sistemi AI. I sistemi di intelligenza artificiale generativa producono significato emergente in modo sistematico e su scala massiccia. Il significato delle risposte di un modello linguistico non è inteso da nessuno — non dagli autori dei testi di addestramento, non dagli ingegneri che hanno progettato il sistema, non dagli annotatori che hanno guidato l'allineamento — ma emerge dall'interazione di tutti questi contributi attraverso l'architettura del modello. Questo rende i sistemi AI il caso limite della semiotica del distribuito: producono significato in modo puramente emergente, senza nessun soggetto intenzionale nemmeno parziale.

La sfida teorica è duplice: capire i meccanismi attraverso cui emerge il significato (una questione di interpretabilità dei modelli che è al centro della ricerca attuale nell'AI) e valutare normativamente questo significato emergente (una questione semiotica ed etica che richiede categorie che la ricerca sull'interpretabilità da sola non può fornire).

VII. Il segno senza soggetto: verso una semiotica post-intenzionale

Le considerazioni sviluppate in questa parte convergono verso la proposta di una semiotica post-intenzionale: una teoria del segno e del significato che non ancori il processo semiotico a un soggetto intenzionale come condizione di possibilità.

La semiotica post-intenzionale non nega che i soggetti intenzionali usino segni con intenzioni comunicative: questo è un fatto che rimane vero per ampia parte dell'uso linguistico. Ma sostiene che l'intenzione del soggetto non è una condizione necessaria né sufficiente per il significato: i segni hanno significato in virtù della loro posizione nelle strutture semiotiche (reti intertestuali, convenzioni linguistiche, pratiche istituzionali) indipendentemente dall'intenzione dei soggetti che li usano, e questa posizione è spesso determinata da processi distribuiti che non hanno un centro intenzionale.

I pilastri di questa semiotica post-intenzionale possono essere enunciati in quattro tesi:

Prima tesi: il significato è posizionale. Il significato di un segno è determinato primariamente dalla sua posizione nelle strutture semiotiche — le reti di relazioni con altri segni, le convenzioni di uso, le pratiche interpretative della comunità — piuttosto che dall'intenzione del soggetto che lo usa. L'intenzione del soggetto contribuisce alla determinazione del significato ma non la determina unilateralmente.

Seconda tesi: il significato è processuale. Il significato di un segno non è una proprietà statica ma il prodotto di un processo di interpretazione che si svolge nel tempo e che può produrre risultati diversi in contesti diversi. Questo processo è distributivo: coinvolge molteplici interpreti, si svolge attraverso molteplici interazioni, e produce convergenze interpretative che non sono riducibili alle interpretazioni individuali dei partecipanti.

Terza tesi: il significato è normativo. Il significato non è semplicemente ciò che i parlanti dicono o intendono, ma ciò che dovrebbero dire e intendere dato il sistema di pratiche semiotiche a cui partecipano. Questa dimensione normativa del significato — che Wittgenstein aveva identificato nella nozione di "uso corretto" — è precisamente la dimensione che la semiotica post-intenzionale deve preservare senza ricorrere all'intenzione del soggetto come fondamento.

Quarta tesi: il significato è emergente. Il significato di un atto linguistico distribuito non è determinabile a priori dalla somma dei significati intesi

dai partecipanti al processo di convergenza: può emergere come proprietà dell'intero processo che non era anticipata da nessuno. Questo non rende il significato emergente arbitrario o incomprensibile, ma richiede strumenti di analisi che vadano oltre l'attribuzione di intenzioni.

Queste quattro tesi non costituiscono ancora una teoria semiotica completa: sono i fondamenti di un programma di ricerca che richiede sviluppo in più direzioni. Ma delineano già la forma di una semiotica del distribuito che può integrare in modo coerente la pragmatica dell'intenzione distribuita con le tradizioni semiotiche più ricche — quella peirciana, quella lotmaniana, quella sociosemiotica — per produrre una teoria del significato adeguata alla complessità degli atti linguistici che governano la vita sociale e politica contemporanea.

Conclusioni sistematiche e programma di ricerca

I. Riepilogo dell'architettura teorica

Il percorso di questa monografia ha attraversato dieci stazioni teoriche che conviene ora raccogliere in una mappa coerente. Non si tratta di un riassunto — di cui il lettore non ha bisogno dopo trecento pagine di argomentazione — ma di un'operazione diversa: identificare le relazioni di dipendenza e supporto reciproco tra le componenti del sistema teorico, in modo da rendere visibile la sua struttura e, con essa, i punti di maggiore solidità e quelli di maggiore fragilità.

L'architettura teorica ha una forma a tre livelli. Il livello fondazionale contiene la critica al principio del soggetto unitario (PSU) e l'analisi della sua genesi storica: questo livello fornisce la motivazione per l'impresa teorica e identifica le aporie della tradizione classica (Austin, Grice, Searle) che la pragmatica distribuita si propone di risolvere. Il livello costruttivo contiene la teoria della funzione di convergenza, con la sua formalizzazione in teoria dei giochi, reti e logica deontica multi-agente, e l'applicazione ai casi empirici: questo livello costituisce il nucleo positivo della proposta teorica. Il livello normativo contiene l'epistemologia del testimone distribuito, l'etica delle strutture di convergenza, la semiotica post-intenzionale, e le conseguenze per la responsabilità, la giustizia

comunicativa e il design istituzionale: questo livello trasforma la proposta teorica da descrizione a normativa.

I tre livelli sono logicamente dipendenti ma non riducibili l'uno all'altro. La critica al PSU fornisce la motivazione ma non determina la forma del livello costruttivo: una critica dello stesso tipo potrebbe in principio portare a proposte teoriche diverse. Il livello costruttivo fornisce la base per il livello normativo ma non lo implica automaticamente: le stesse strutture descrittive possono supportare standard normativi diversi a seconda dei valori che si ritiene opportuno privilegiare. La coerenza del sistema dipende dall'esplicitazione delle relazioni tra questi livelli — operazione che questa parte conclusiva si propone di completare.

II. La funzione di convergenza: bilancio e limiti

La funzione di convergenza è il concetto centrale della pragmatica distribuita e merita un bilancio critico che ne identifichi sia i contributi sia i limiti.

Contributi. La funzione di convergenza consente di: (a) descrivere come sistemi distribuiti di complessità arbitraria producano atti linguistici con forza illocutoria determinata, senza presupporre un soggetto intenzionale unitario; (b) distinguere tra condizioni strutturali e sostanziali di felicità, catturando fenomeni come la validità formale senza fondatezza epistemica; (c) articolare la responsabilità per gli atti linguistici distribuiti in tre livelli — strutturale, epistemica, per design — che superano le limitazioni delle categorie giuridiche tradizionali; (d) fornire misure quantitative di qualità della convergenza (adeguatezza epistemica, equità, robustezza) che consentono confronti tra strutture diverse; (e) collegare la pragmatica degli atti linguistici alle teorie dell'organizzazione, della decisione collettiva e dei sistemi complessi in modo produttivo e non riducibile a mere analogie.

Limiti. La funzione di convergenza presenta limitazioni che devono essere riconosciute per non sopravvalutare la proposta. Il primo limite è quello della circolarità parziale: la funzione di convergenza è definita come il processo che trasforma la molteplicità in output con forza illocutoria determinata, ma la determinazione della forza illocutoria dipende a sua volta — come si è visto nella terza parte — dalla struttura di convergenza. Questa circolarità non è necessariamente viziosa, ma

richiede che il framework sia sviluppato in modo da non presupporre ciò che vuole spiegare.

Il secondo limite è quello della granularità: la funzione di convergenza descrive le proprietà dell'output del sistema, ma non descrive i meccanismi micro-livello attraverso cui le preferenze e i contributi degli agenti individuali si integrano nel processo. Il livello di analisi meso-sociale della funzione di convergenza deve essere integrato con analisi micro-sociali (come si comportano individualmente gli agenti nel processo) e macro-sociali (come le strutture di convergenza si inseriscono nei sistemi sociali più ampi) per produrre un quadro completo.

Il terzo limite è quello della misurabilità: le tre misure di qualità della convergenza proposte nella sesta parte sono concettualmente precise ma operazionalmente difficili da applicare ai casi reali. La misura di adeguatezza epistemica, in particolare, richiede di identificare "la totalità delle informazioni rilevanti disponibili nel sistema" — un insieme che è raramente delimitato con precisione nei casi reali. L'operazionalizzazione di queste misure richiede lavoro empirico e metodologico che questa monografia non può condurre.

III. Relazioni con le tradizioni teoriche: convergenze e divergenze

Una proposta teorica si definisce anche per le sue relazioni con le tradizioni esistenti. È utile esplicitare queste relazioni in modo da rendere visibile il contributo originale della pragmatica distribuita rispetto a ciò che era già disponibile.

Rispetto alla pragmatica classica (Austin, Grice, Searle), la pragmatica distribuita non è una sostituzione ma un'estensione critica. Mantiene l'impianto fondamentale della teoria degli atti linguistici — la distinzione tra atti illocutori e perlocutori, la nozione di condizioni di felicità, l'analisi della forza illocutoria — ma rimuove il presupposto del soggetto unitario come condizione di applicabilità. Il risultato non invalida la pragmatica classica per i suoi casi d'elezione (la comunicazione interpersonale faccia a faccia tra adulti psicologicamente integrati) ma la subordina a un quadro più generale in cui questi casi sono casi limite.

Rispetto alla teoria delle organizzazioni (Simon, Cyert e March, Weick), la pragmatica distribuita applica alle organizzazioni categorie che quella tradizione non aveva sviluppato sistematicamente: la forza illocutoria, le condizioni di felicità, la responsabilità per atti linguistici. Il contributo è bidirezionale: la teoria delle organizzazioni fornisce modelli empirici di come i sistemi distribuiti prendono decisioni; la pragmatica distribuita fornisce categorie normative per valutare gli atti comunicativi che quelle decisioni producono.

Rispetto alla filosofia del diritto (Hart, Dworkin, Alexy), la pragmatica distribuita offre una base teorica per la comprensione degli atti linguistici giuridici — sentenze, leggi, contratti — che non dipende da finzioni di intenzione unitaria del "legislatore" o del "giudice". La teoria dell'interpretazione giuridica, in particolare, può beneficiare significativamente della categoria di ambiguità intenzionale distribuita come alternativa all'interpretazione intenzionalista e come strumento per analizzare i conflitti interpretativi che emergono da processi legislativi distributivi.

Rispetto alla filosofia politica deliberativista (Habermas, Rawls, Cohen), la pragmatica distribuita offre una critica costruttiva: non all'ideale deliberativo in sé ma al modo in cui viene concettualizzato il soggetto della deliberazione. Allargare la teoria della legittimità democratica alle strutture interne di convergenza dei soggetti politici collettivi — come suggerito nella quarta parte — è un'estensione del programma deliberativista che il framework della pragmatica distribuita rende possibile.

Rispetto all'etica dell'intelligenza artificiale, la pragmatica distribuita offre strumenti che il dibattito corrente manca: non solo principi etici per il comportamento dei sistemi AI, ma categorie analitiche per valutare le strutture di convergenza che li producono. La distinzione tra adeguatezza epistemica, equità e robustezza della convergenza offre una grammatica più precisa per il dibattito sul fairness, sulla trasparenza e sulla sicurezza dei sistemi AI rispetto alle nozioni generiche attualmente dominanti nel dibattito.

IV. Questioni aperte: una mappa

Il percorso di questa monografia ha aperto più questioni di quante ne abbia risolto. Questa è, come si notava nella conclusione originale, un segno di fecondità teorica piuttosto che di incompletezza. Ma la fecondità è produttiva solo se le questioni aperte sono identificate con precisione sufficiente da orientare la ricerca futura. Si propone una mappa delle questioni aperte organizzata in quattro cluster.

Cluster 1: Formalizzazione e misura. Le strutture formali proposte nella sesta parte — giochi di convergenza, reti di influenza, logica DMAL — sono abbozzi di un programma di formalizzazione che richiede sviluppo sistematico. Le questioni aperte principali riguardano: l'assiomatizzazione completa della logica DMAL con prove di completezza e decidibilità; la costruzione di misure operazionalizzabili di adeguatezza epistemica, equità e robustezza che possano essere applicate a casi reali con dati disponibili; l'integrazione delle misure di qualità della convergenza in un framework di welfare comunicativo che consenta comparazioni tra strutture istituzionali diverse.

Cluster 2: Fondamenti cognitivi. La tesi che il singolo soggetto cognitivo sia già un sistema distribuito — centrale nella seconda parte — è supportata da evidenze neuroscientifiche e psicologiche, ma non è stata sviluppata in una teoria cognitiva sistematica degli atti linguistici. Le questioni aperte riguardano: il modo in cui i processi sub-personali di produzione linguistica si integrano nella produzione di atti comunicativi con struttura intenzionale riconoscibile; il ruolo della coscienza come processo di integrazione e non come centro intenzionale; le implicazioni per la patologia della comunicazione (afasie, dissociazioni, disturbi della comunicazione) di una teoria che non presuppone l'unitarietà del soggetto.

Cluster 3: Applicazioni giuridiche. Le categorie della responsabilità strutturale, della responsabilità epistemica distribuita e della responsabilità per design della convergenza hanno bisogno di sviluppo dottrinale prima di potere essere applicate nei sistemi giuridici reali. Le questioni aperte riguardano: il modo in cui la responsabilità strutturale può essere incorporata nei sistemi di diritto civile e penale degli enti senza produrre incongruenze con i principi fondamentali di questi sistemi; i criteri per determinare quando una struttura di convergenza è "adeguata" in senso normativo sufficientemente preciso da poter fondare giudizi di responsabilità; l'applicazione alle giurisdizioni che hanno già sviluppato forme di responsabilità degli enti (USA, Francia, Italia) rispetto a quelle che non le hanno.

Cluster 4: Sistemi AI e governance. Le questioni legate ai sistemi di intelligenza artificiale generativa sono probabilmente le più urgenti dal punto di vista applicativo. Le questioni aperte principali riguardano: lo sviluppo di standard tecnici per la trasparenza delle strutture di convergenza dei sistemi AI (quali informazioni sul corpus di addestramento, sulle tecniche di allineamento, sui meccanismi di governance devono essere rese pubbliche e in quale forma); i criteri per attribuire responsabilità per gli output dei sistemi AI tra produttori, deployers e utenti in modo coerente con il framework della responsabilità distribuita; il disegno di meccanismi di governance dei sistemi AI che rispettino i principi dell'epistemologia normativa del distribuito (mobilizzazione, diversità, tracciabilità, revisione, protezione del dissenso).

V. Un programma di ricerca interdisciplinare

Le questioni aperte identificate nella sezione precedente non possono essere affrontate da una singola disciplina. La pragmatica dell'intenzione distribuita è un progetto interdisciplinare per necessità: richiede la collaborazione tra filosofia del linguaggio, linguistica, teoria delle organizzazioni, informatica, diritto, scienze politiche e — non ultimo — neuroscienze cognitive e psicologia. Questa interdisciplinarità non è un lusso ma una condizione di adeguatezza teorica: la complessità del fenomeno non può essere catturata dai concetti di nessuna singola disciplina.

Il programma di ricerca che questa monografia suggerisce ha tre linee principali.

Linea A: Fondamenti teorici. Sviluppo della formalizzazione della funzione di convergenza; assiomatizzazione della logica DMAL; costruzione di modelli di agent-based simulation per testare le predizioni della teoria su strutture di convergenza artificiali; integrazione con le teorie cognitive della produzione linguistica.

Linea B: Ricerca empirica. Analisi empirica delle strutture di convergenza in sistemi istituzionali diversi (parlamenti, tribunali, organizzazioni internazionali, aziende); sviluppo di strumenti di misura dell'adeguatezza epistemica, dell'equità e della robustezza applicabili a

dati reali; studi longitudinali sulla stabilità degli atti linguistici distribuiti in relazione alle caratteristiche della convergenza che li ha prodotti.

Linea C: Sviluppo normativo. Elaborazione dottrinale delle categorie di responsabilità strutturale e per design della convergenza nei sistemi giuridici esistenti; sviluppo di standard di governance per i sistemi AI basati sui principi della pragmatica distribuita; proposta di riforme istituzionali per la democratizzazione delle strutture di convergenza nei sistemi politici; elaborazione di codici etici per la comunicazione istituzionale basati sui principi della giustizia comunicativa.

VI. La posta in gioco: perché la pragmatica distribuita è necessaria ora

Una teoria filosofica non viene valutata solo per la sua coerenza interna o per la sua eleganza formale, ma anche per la sua rilevanza rispetto ai problemi del tempo in cui viene proposta. La pragmatica dell'intenzione distribuita è una proposta teorica che risponde a problemi specifici del momento storico: problemi che la tradizione classica non era attrezzata a vedere, perché erano oscurati dal PSU, e che sono diventati urgenti proprio nell'epoca in cui i sistemi distribuiti di produzione di atti linguistici hanno raggiunto una scala e una complessità senza precedenti storici.

La crisi della democrazia liberale è in parte una crisi della comunicazione politica: l'incapacità di distinguere tra informazione e disinformazione, tra deliberazione genuina e manipolazione strutturale, tra legittimità formale e legittimità sostanziale. La pragmatica distribuita offre strumenti per diagnosticare questa crisi in modo più preciso di quanto consentano le categorie tradizionali: non come crisi della menzogna individuale ma come crisi delle strutture di convergenza che producono il discorso pubblico.

La diffusione su scala globale dei sistemi di intelligenza artificiale generativa ha reso urgente la questione di come attribuire responsabilità e significato agli atti linguistici prodotti da sistemi senza soggetto intenzionale. La pragmatica distribuita offre gli strumenti concettuali per affrontare questa questione senza cadere nella trappola della personificazione (attribuire ai sistemi AI intenzioni e responsabilità di

tipo umano) né in quella del nichilismo pragmatico (negare che questi sistemi producano atti linguistici con forza normativa reale).

La crisi della governance globale — l'incapacità delle istituzioni internazionali di produrre atti linguistici con forza normativa adeguata alle sfide collettive (cambiamento climatico, pandemie, regolazione della tecnologia) — è in parte una crisi delle strutture di convergenza di quelle istituzioni. La pragmatica distribuita offre una diagnosi precisa: strutture di convergenza che non soddisfano i principi di inclusività, adeguatezza epistemica e robustezza producono sistematicamente atti linguistici con forza illocutoria insufficiente. E offre anche indicazioni per il disegno istituzionale: non mere aspirazioni all'inclusività, ma principi tecnici per la costruzione di strutture di convergenza che massimizzino la qualità epistemica e la legittimità democratica degli output.

VII. L'unità della proposta: la pragmatica come scienza dell'architettura comunicativa

Si era proposto, nella conclusione della prima versione di questa monografia, di intendere la pragmatica come "scienza dell'architettura comunicativa". Questa definizione può ora essere precisata alla luce del percorso compiuto nelle parti aggiuntive.

La pragmatica come scienza dell'architettura comunicativa studia i modi in cui le strutture attraverso cui i sistemi distribuiti producono atti linguistici determinano: (a) la forza illocutoria degli atti — quale tipo di azione è compiuta e con quale intensità normativa; (b) la qualità epistemica degli atti — in quale misura rispecchiano le informazioni rilevanti disponibili nel sistema; (c) la giustizia comunicativa degli atti — in quale misura riflettono in modo equo le prospettive e gli interessi di tutti gli agenti concerniti; (d) la responsabilità per gli atti — come è distribuita tra i partecipanti al processo di convergenza e tra i progettisti delle strutture.

Questa definizione recupera, trasformandole, le tre funzioni che l'intenzione svolgeva nella teoria classica — individuazione dell'atto, condizioni di felicità, principio di responsabilità — e le ridistribuisce tra la struttura della convergenza, la posizione istituzionale dell'output, e i principi normativi che governano il design. Il soggetto unitario non scompare del tutto: rimane un caso limite importante, e la sua psicologia

individuale continua ad essere rilevante per molti tipi di atti linguistici. Ma non è più il fondamento insostituibile su cui regge l'intera architettura: è un elemento tra altri in un sistema teorico più ricco e più adeguato alla complessità del linguaggio reale.

Il linguaggio reale è prodotto da sistemi distribuiti. Una pragmatica adeguata a questo linguaggio deve essere una scienza delle strutture attraverso cui la molteplicità si converte in forza illocutoria, e dei principi normativi che dovrebbero governare questa conversione. Questa monografia ha cercato di mostrare che una tale scienza è possibile, di indicarne le linee principali, e di dimostrarne la necessità. Il suo completamento è un'impresa collettiva — che richiede, con una certa ironia, il tipo di convergenza distribuita che è stata al centro della sua analisi.

Appendice A

Neuroscienze cognitive e distribuzione sub-personale: approfondimenti

A.1 Il paradigma Libet e le sue implicazioni per la teoria dell'intenzione

Gli esperimenti condotti da Benjamin Libet negli anni Ottanta del Novecento rappresentano uno degli snodi empirici più discussi nel dibattito contemporaneo sul libero arbitrio e sulla struttura dell'intenzione. La loro rilevanza per la pragmatica distribuita è diretta e merita una trattazione più dettagliata di quella offerta nel corpo principale della monografia.

Il protocollo sperimentale di Libet chiedeva ai partecipanti di compiere un movimento volontario del polso in un momento di loro scelta, annotando il momento in cui avvertivano l'impulso cosciente a muoversi attraverso la posizione di un indicatore su un quadrante rotante. L'attività cerebrale veniva registrata tramite EEG. Il risultato principale — replicato in numerose varianti nei decenni successivi — mostrava che il cosiddetto Readiness Potential (Bereitschaftspotential), un segnale elettrico che riflette la preparazione motoria del cervello, precedeva il momento dell'intenzione cosciente riferita dai partecipanti di circa 550 millisecondi, e precedeva il movimento effettivo di circa 800 millisecondi.

L'interpretazione standard — che l'intenzione cosciente non "causa" il movimento ma arriva troppo tardi per farlo — è stata oggetto di controversia metodologica intensa. Critiche legittime riguardano la precisione della stima del momento dell'intenzione cosciente (basata su auto-report in tempo reale), la natura del Readiness Potential come indicatore di preparazione piuttosto che di decisione, e l'artificialità della situazione sperimentale rispetto alle intenzioni comunicative ordinarie.

Tuttavia, il paradigma Libet — considerato insieme alle ricerche successive di Haynes (2008), che hanno mostrato capacità predittive del comportamento fino a dieci secondi prima della consapevolezza soggettiva, e di Soon et al. (2008) su decisioni bi-alternative — suggerisce una conclusione più moderata ma ancora rilevante per la pragmatica: la coscienza dell'intenzione non è necessariamente il punto di origine causale dell'atto, ma può essere una rappresentazione post-hoc (o, più cautamente, a latenza variabile) di processi che si svolgono a livello sub-personale. Questa conclusione non elimina il ruolo della coscienza nella regolazione del comportamento — Libet stesso ha proposto che la coscienza eserciti una funzione di "veto" sulle preparazioni motorie già avviate — ma modifica radicalmente la struttura della relazione tra intenzione cosciente e atto.

Per la pragmatica distribuita, la rilevanza è questa: se anche nel caso del singolo agente individuale l'intenzione cosciente è il prodotto di processi sub-personali distribuiti piuttosto che il punto di origine dell'atto, allora la distinzione tra soggetti individuali "con intenzione unitaria" e sistemi distribuiti "senza intenzione" è una distinzione di grado piuttosto che di tipo. Il singolo soggetto è già un sistema distribuito che produce una rappresentazione integrata dell'intenzione come output di processi non accessibili alla coscienza.

A.2 Gazzaniga e l'interprete: narrazione come convergenza post-hoc

Michael Gazzaniga, negli studi che si estendono per oltre quattro decenni con pazienti sottoposti a callosotomia (sezione del corpo calloso per il trattamento dell'epilessia intrattabile), ha prodotto evidenze che illuminano la struttura distributiva del soggetto cognitivo con una chiarezza drammatica non disponibile attraverso altre metodologie.

I risultati classici sono noti: presentando stimoli diversi ai due emisferi attraverso presentazioni tachistoscopiche nel campo visivo appropriato, è possibile dimostrare che i due emisferi operano come agenti

semi-indipendenti con credenze, preferenze e capacità comunicative diverse. Ciò che è meno noto, ma teoricamente più ricco, è il meccanismo che Gazzaniga ha denominato "interprete dell'emisfero sinistro": un processo che produce narrazioni causali coerenti a partire da comportamenti generati da processi che l'emisfero sinistro non controlla né ha accesso diretto.

Nell'esperimento paradigmatico, il paziente riceve simultaneamente l'istruzione "cammina" nell'emisfero destro (che non ha capacità verbali) e una diversa istruzione nell'emisfero sinistro. Il paziente obbedisce all'istruzione ricevuta dall'emisfero destro (alza e comincia a camminare). Quando gli viene chiesto perché si è alzato, l'emisfero sinistro — che non ha accesso all'istruzione ricevuta dall'altro emisfero — produce immediatamente una spiegazione causale plausibile: "Volevo andare a prendere qualcosa da bere."

Questa narrazione post-hoc non è una bugia: è il prodotto di un meccanismo interpretativo che genera spiegazioni causalmente plausibili per comportamenti il cui reale antecedente causale non è accessibile. Gazzaniga ha chiamato questo meccanismo "interprete" proprio perché la sua funzione è ermeneutica: costruisce significato narrativo a partire da dati comportamentali, esattamente come un interprete costruisce senso a partire da un testo che non ha prodotto.

Le implicazioni per la pragmatica sono profonde. Se il soggetto individuale produce narrazioni intenzionali post-hoc a partire da comportamenti verbali generati da processi sub-personali, allora la struttura dell'intenzione comunicativa non è quella di un piano che precede e causa l'atto, ma di una narrazione che lo segue e lo razionalizza. La distinzione tra intenzione genuina e razionalizzazione post-hoc — cruciale per la teoria classica degli atti linguistici, che considera l'intenzione come condizione di felicità dell'atto — diventa molto più sfumata di quanto si supponga.

Per la pragmatica distribuita, l'interprete di Gazzaniga è un modello di come ogni sistema cognitivo — individuale o collettivo — produce rappresentazioni della propria intenzione come output di processi di integrazione e narrazione, piuttosto che come punto di origine causale degli atti. La funzione di convergenza, a livello sub-personale, è precisamente questo processo di integrazione che produce una rappresentazione coerente dell'intenzione a partire da processi distribuiti

che non avevano necessariamente una struttura intenzionale unitaria prima della convergenza.

A.3 La cognizione estesa: Clark e Chalmers

La teoria della mente estesa, proposta da Andy Clark e David Chalmers nel loro influente articolo del 1998, introduce il principio di parità: se uno strumento esterno al corpo svolge una funzione che, se svolta all'interno del cranio, sarebbe classificata come cognitiva, allora anche lo strumento esterno fa parte del processo cognitivo. Il caso paradigmatico è quello di Otto, un paziente con compromissione della memoria episodica che porta sempre con sé un taccuino su cui annota informazioni importanti. Clark e Chalmers sostengono che il taccuino fa parte del sistema cognitivo di Otto nello stesso senso in cui la memoria biologica fa parte del sistema cognitivo di Inga.

Questa proposta ha implicazioni dirette per la pragmatica distribuita. Se la cognizione può estendersi al di là dei confini biologici del soggetto individuale per includere artefatti e ambienti, allora la produzione di atti linguistici — che dipende da risorse cognitive — è anch'essa potenzialmente estesa: non è localizzata esclusivamente nel cervello del parlante ma distribuita attraverso i sistemi cognitivi estesi che il parlante usa.

Nel contesto istituzionale, questa prospettiva è particolarmente rilevante: le organizzazioni costruiscono sistemi di memoria esterna (archivi, banche dati, procedure codificate), sistemi di elaborazione esterna (protocolli decisionali, sistemi informativi, procedure di consultazione), e sistemi di comunicazione esterna (canali ufficiali, stili comunicativi istituzionali) che fanno parte, nel senso di Clark e Chalmers, del sistema cognitivo esteso dell'organizzazione. La funzione di convergenza dell'organizzazione non è solo il processo attraverso cui le menti individuali dei suoi agenti si coordinano: è il processo attraverso cui l'intero sistema cognitivo esteso dell'organizzazione produce output comunicativi.

Questa prospettiva suggerisce un'estensione del framework della pragmatica distribuita verso una pragmatica degli artefatti cognitivi: un'analisi di come le strutture di memoria, elaborazione e comunicazione esternalizzate influenzino le proprietà degli atti linguistici prodotti dai sistemi che le utilizzano. I sistemi informativi, i protocolli di decisione, le strutture di governance documentale — tutti questi artefatti fanno parte

del sistema cognitivo esteso delle organizzazioni e influenzano sistematicamente la funzione di convergenza.

A.4 La predictive processing e la comunicazione come inferenza attiva

La teoria del predictive processing (o active inference), sviluppata da Karl Friston e collaboratori come framework generale per la cognizione, offre un'ulteriore prospettiva rilevante per la pragmatica distribuita. In questa teoria, il cervello è un sistema che genera costantemente predizioni sul proprio ambiente (e su se stesso) e aggiorna queste predizioni minimizzando l'errore di predizione attraverso due meccanismi complementari: l'aggiornamento delle predizioni alla luce dei dati sensoriali (inferenza perceptiva) e l'azione sull'ambiente per far corrispondere i dati sensoriali alle predizioni (inferenza attiva).

La comunicazione linguistica, in questa prospettiva, è una forma di inferenza attiva: il parlante produce enunciati che agiscono sull'ambiente (modifica le credenze dell'ascoltatore) in modo da ridurre l'errore di predizione rispetto a un modello interno del suo stato desiderato del mondo. Ma questo processo non richiede che il parlante abbia un'intenzione cosciente esplicita: le predizioni sono generate a livelli gerarchici multipli del sistema cerebrale, la maggior parte dei quali non è accessibile alla coscienza.

Per la pragmatica distribuita, il predictive processing offre un modello di come la produzione linguistica sia già, a livello sub-personale, un processo distributivo: diversi livelli gerarchici del sistema cerebrale generano predizioni a diversi livelli di astrazione (fonologico, morfologico, pragmatico, semantico) e le coordinate in un output linguistico che emerge dalla minimizzazione dell'errore attraverso questi livelli. Il soggetto "che intende comunicare" è il risultato di questo processo di integrazione gerarchica distribuita, non il suo punto di origine.

Questa prospettiva apre anche una domanda interessante per la comunicazione istituzionale: come si realizza il predictive processing a livello di sistemi distribuiti? Esiste un analogo funzionale della minimizzazione dell'errore di predizione nei processi organizzativi di produzione comunicativa? La risposta tentativa della pragmatica distribuita è che la funzione di convergenza svolge questo ruolo: è il processo attraverso cui il sistema organizzativo riduce la discrepanza tra

le predizioni dei diversi sotto-sistemi e produce un output comunicativo che rappresenta il punto di minimo errore di predizione aggregato.

A.5 Implicazioni per le patologie della comunicazione istituzionale

Le neuroscienze cognitive offrono, oltre ai modelli positivi di come funziona la cognizione distribuita, modelli delle sue patologie che sono illuminanti per la pragmatica distribuita. Le dissociazioni prodotte da lesioni cerebrali o da disturbi dello sviluppo mostrano cosa accade quando i meccanismi di integrazione distributiva falliscono: comportamenti verbali non coordinati, narrazioni incoerenti, incapacità di integrare informazioni da fonti diverse.

Le organizzazioni presentano patologie funzionalmente analoghe quando i loro meccanismi di convergenza falliscono: comunicazioni interne contraddittorie che producono messaggi esterni incoerenti; incapacità di integrare informazioni da unità diverse; narrazioni organizzative ufficiali che divergono sistematicamente dai comportamenti effettivi; blocchi nei canali di comunicazione verticale che producono il fenomeno dell'ignoranza organizzativa (organizational blindness) analogo alla negligenza percettiva nei pazienti con lesioni dell'attenzione.

Studiare queste patologie organizzative attraverso la lente delle neuroscienze cognitive — non come analogia metaforica ma come parallelo strutturale tra sistemi di integrazione distributiva a livelli diversi — è uno dei programmi di ricerca più promettenti che la pragmatica distribuita apre. La tassonomia delle patologie della comunicazione istituzionale che una tale ricerca potrebbe produrre avrebbe valore sia descrittivo (classificare i tipi di guasto delle strutture di convergenza) sia normativo (identificare i meccanismi di integrazione la cui assenza produce ciascun tipo di patologia).

Appendice B

Diritto comparato della responsabilità degli enti per atti linguistici: schema

B.1 Il problema della responsabilità degli enti nel diritto contemporaneo

I sistemi giuridici moderni affrontano in modo eterogeneo il problema della responsabilità degli enti per atti comunicativi che producono danno. Questa eterogeneità riflette differenze nelle tradizioni giuridiche, nelle strutture economiche e nelle scelte di politica del diritto, ma anche — come si argomenterà — una mancanza di strumenti teorici adeguati per concettualizzare la responsabilità distribuita. L'analisi comparata dei sistemi principali consente di identificare le soluzioni esistenti, i loro limiti, e le direzioni in cui il framework della pragmatica distribuita potrebbe contribuire al loro sviluppo.

B.2 Il sistema statunitense: responsabilità per comunicazioni fraudolente e due process

Il diritto statunitense ha sviluppato uno dei sistemi più articolati di responsabilità degli enti per atti comunicativi, principalmente attraverso il Securities Exchange Act (1934) e la sua applicazione da parte della SEC, il Foreign Corrupt Practices Act (1977), e il common law della fraud e misrepresentation.

La Rule 10b-5 della SEC — che proibisce qualunque dispositivo, schema o artificio per frodare, qualunque dichiarazione falsa di fatto materiale, o qualunque omissione che renda le dichiarazioni fatte ingannevoli — ha generato una giurisprudenza vastissima sulla responsabilità delle corporazioni per dichiarazioni false o fuorvianti ai mercati finanziari. Il caso Basic Inc. v. Levinson (1988) della Corte Suprema ha stabilito la dottrina della fraud-on-the-market theory: in mercati efficienti, gli investitori hanno diritto di presumere che il prezzo di mercato rifletta tutte le informazioni pubblicamente disponibili, incluse le dichiarazioni dell'emittente; le dichiarazioni false che alterano questo prezzo producono responsabilità verso gli investitori anche senza prova di reliance individuale.

Dal punto di vista della pragmatica distribuita, il sistema americano ha sviluppato un meccanismo di responsabilità che è funzionalmente equivalente alla responsabilità strutturale proposta nel testo principale: la corporazione risponde per gli effetti delle sue dichiarazioni sui mercati indipendentemente dalla prova di intenzione fraudolenta di specifici individui. Ma il fondamento teorico rimane insoddisfacente: il diritto americano costruisce questa responsabilità attraverso la finzione della corporate intent piuttosto che attraverso una teoria della responsabilità strutturale genuina.

Il Dodd-Frank Wall Street Reform Act (2010) ha introdotto meccanismi di whistleblower che indirettamente riconoscono la struttura distributiva della produzione di informazioni finanziarie: il whistleblower è l'agente che ha accesso a informazioni rilevanti che non sono state incorporate nel processo di convergenza ufficiale, e il sistema lo protegge e lo incentiva precisamente perché il suo contributo corregge una distorsione strutturale della convergenza.

B.3 Il sistema italiano: il d.lgs. 231/2001 e i suoi limiti

Il decreto legislativo 231/2001 ha introdotto in Italia la responsabilità amministrativa degli enti per i reati commessi nel loro interesse o a loro vantaggio da soggetti in posizione apicale o sottoposti alla loro direzione. Questo sistema — modellato su esperienze internazionali e adottato in risposta alle direttive europee sulla responsabilità degli enti — rappresenta un passo significativo verso la responsabilità delle organizzazioni come tali, piuttosto che come aggregati di individui responsabili.

Le fattispecie rilevanti per la responsabilità degli enti per atti linguistici includono i reati societari (false comunicazioni sociali, falso in prospetto), i reati tributari, i reati contro la pubblica amministrazione (corruzione, concussione, spesso commessi attraverso atti comunicativi), e i reati di mercato (aggiotaggio, insider trading).

Il sistema 231 esonera l'ente dalla responsabilità se dimostra di aver adottato ed efficacemente attuato, prima della commissione del reato, un modello di organizzazione, gestione e controllo idoneo a prevenire reati della specie di quello verificatosi. Questa clausola — il cosiddetto modello organizzativo esimente — è l'elemento più interessante per la pragmatica distribuita: riconosce implicitamente che la struttura organizzativa (la struttura di convergenza, nel nostro linguaggio) è rilevante per la responsabilità dell'ente, e che la progettazione di strutture adeguate è un dovere dell'ente stesso.

Tuttavia, il sistema 231 presenta limiti significativi che il framework della pragmatica distribuita aiuta a identificare. Il primo è che richiede l'identificazione di un reato commesso da una persona fisica che sia "autore" dell'illecito: la responsabilità dell'ente è derivata da quella dell'individuo. Questo rende il sistema inadeguato per i casi di responsabilità strutturale pura, in cui il danno è prodotto dalla struttura

di convergenza senza che nessun individuo abbia commesso un illecito identificabile.

Il secondo limite è che il modello organizzativo esimente è concepito in termini di prevenzione dei reati individuali, non di progettazione di strutture di convergenza epistemicamente adeguate. Un ente potrebbe avere un modello 231 formalmente impeccabile che non produce tuttavia strutture di convergenza in grado di mobilitare le informazioni rilevanti per prevenire atti comunicativi dannosi che non configurano reati penali (dichiarazioni fuorvianti non fraudolente, omissioni non dolosi).

B.4 Il sistema europeo: responsabilità per prodotti difettosi e comunicazione AI

Il diritto europeo ha sviluppato, attraverso decenni di direttive e regolamenti, un sistema di responsabilità degli enti che si estende progressivamente verso la responsabilità per la struttura dei processi produttivi piuttosto che per i difetti dei singoli prodotti. La Direttiva sulla responsabilità per danno da prodotti difettosi (85/374/CEE), riformata dalla proposta di regolamento del 2022, si è progressivamente allargata a includere sistemi software e dati come "prodotti" che possono essere difettosi.

L'AI Liability Directive (proposta 2022, ancora in corso di adozione al momento di chiusura del presente lavoro) rappresenta il tentativo più diretto di affrontare il problema della responsabilità per gli output dei sistemi AI. La proposta introduce un meccanismo di presunzione di nesso causale: se il sistema AI ha operato in modo che avrebbe prodotto il danno allegato, e se il deployer non ha rispettato gli obblighi di diligenza previsti dall'AI Act, si presume il nesso causale tra il comportamento del sistema e il danno.

Questo meccanismo è interessante perché sposta il focus dalla prova dell'intenzione o della colpa individuale alla prova del rispetto degli standard di progettazione e deployment del sistema — che è precisamente il tipo di responsabilità per design della convergenza che la pragmatica distribuita propone. Tuttavia, la proposta rimane vincolata al paradigma della responsabilità per danno individuale e non sviluppa pienamente gli strumenti per la responsabilità per danni sistemici prodotti da strutture di convergenza inadeguate senza un danno individuale identificabile.

B.5 Verso una convergenza giuridica internazionale: i principi di responsabilità distribuita

L'analisi comparata suggerisce che i sistemi giuridici stiano convergendo — in modo non coordinato ma parallelo — verso forme di responsabilità che si avvicinano alla responsabilità strutturale proposta dalla pragmatica distribuita: responsabilità degli enti per le caratteristiche dei loro processi organizzativi piuttosto che per i comportamenti individuali dei loro membri. Questa convergenza è guidata dalla pressione di fenomeni empirici — crisi finanziarie sistemiche, scandali aziendali su larga scala, danni da sistemi AI — che le categorie tradizionali non riescono a gestire adeguatamente.

Il contributo che la pragmatica distribuita può offrire a questa convergenza è teorico: fornire i concetti che i sistemi giuridici stanno cercando empiricamente attraverso soluzioni ad hoc. I principi di responsabilità distribuita elaborati nel corpo della monografia — responsabilità strutturale, responsabilità epistemica distribuita, responsabilità per design della convergenza — offrono una grammatica precisa per lo sviluppo di questo diritto emergente.

Appendice C

Tavole dei modelli formali

C.1 Tavola sinottica della logica DMAL

La tavola seguente riassume la sintassi e la semantica dei principali operatori della Deontic Multi-Agent Logic (DMAL) introdotta nella Parte Sesta, con annotazioni sull'interpretazione pragmatica di ciascun operatore.

Operatore	Notazione	Lettura	Interpretazione pragmatica
Conoscenza individuale	$K_i \varphi$	"l'agente i sa che φ"	accesso epistemico individuale all'informazione rilevante per l'atto

Credenza individuale	B_i φ	"l'agente i crede che φ"	stato doxastico individuale, può divergere da K_i
Intenzione individuale	I_i φ	"l'agente i intende che φ"	stato pro-attivo individuale verso l'output φ
Obbligo individuale	O_i φ	"l'agente i è obbligato a che φ"	vincolo deontologico individuale nel processo
Permesso individuale	P_i φ	"all'agente i è permesso che φ"	spazio d'azione deonticamente ammissibile
Divieto individuale	F_i φ	"all'agente i è vietato che φ"	vincolo negativo nel processo
Conoscenza collettiva	K_C φ	"la coalizione C sa collettivamente che φ"	informazione disponibile nel sistema distribuito
Obbligo collettivo	O_C φ	"la coalizione C ha l'obbligo collettivo che φ"	vincolo deontologico della struttura
Asserzione	Assert(i, φ)	"l'agente i asserisce φ"	atto illocutorio assertivo individuale
Promessa	Promise(i, φ)	"l'agente i promette φ"	atto illocutorio commissivo individuale
Convergenza	Conv(C, α)	"la coalizione C converge su α"	output della funzione di convergenza
Validità strutturale	Valid_P(C, α)	"α è strutturalmente valido per C secondo P"	conformità procedurale dell'atto distribuito
Felicità sostanziale	SubstFel(α)	"α è sostanzialmente felice"	fondatezza epistemica dell'atto distribuito

Assiomi fondamentali della DMAL:

(A1) Distribuzione della conoscenza collettiva:

$K_C \varphi \rightarrow \exists\, i \in C : K_i \varphi$

La conoscenza collettiva presuppone che almeno un membro della coalizione possieda l'informazione individualmente. Non richiede che tutti la posseggano.

(A2) Convergenza e validità strutturale:

$Conv(C, \alpha) \wedge Proc_P(C, \alpha) \rightarrow Valid_P(C, \alpha)$

Se la coalizione C converge su α seguendo la procedura P, l'atto è strutturalmente valido rispetto a P.

(A3) Dissociabilità di validità strutturale e felicità sostanziale:

$Valid_P(C, \alpha) \not\Rightarrow SubstFel(\alpha)$

La validità strutturale non implica la felicità sostanziale. Questo assioma formalizza l'indipendenza logica tra le due dimensioni di adeguatezza degli atti distribuiti, lasciando aperta la possibilità di dissociazione empirica.

(A4) Condizione necessaria per la responsabilità epistemica individuale:

$Resp(i, \alpha) \leftarrow [K_i\ contrib(i, \alpha) \vee B_i\ contrib(i, \alpha)] \wedge [P_i \neg\alpha \vee O_i\ signal(\neg SubstFel(\alpha))]$

L'agente i è (parzialmente) responsabile dell'atto distribuito α se: (i) sapeva o credeva di contribuire causalmente alla sua produzione, e (ii) aveva o il potere di prevenire α oppure l'obbligo di segnalare al sistema la mancata soddisfazione delle condizioni sostanziali di felicità.

(A5) Monotonia della conoscenza collettiva rispetto all'inclusione di coalizioni:

$C \subseteq C' \wedge K_C \varphi \rightarrow K_C' \varphi$

Se la coalizione C sa collettivamente φ, ogni coalizione più ampia che include C sa anch'essa collettivamente φ. L'aggiunta di agenti non distrugge la conoscenza collettiva già posseduta.

(A6) Validità strutturale e responsabilità collettiva dell'ente:

$\text{Valid_P}(N, \alpha) \wedge \text{Control}(N, P) \rightarrow \text{RespStruct}(N, \alpha)$

L'ente N ha responsabilità strutturale per α se l'atto è strutturalmente valido secondo una procedura P che era sotto il controllo normativo dell'ente stesso.

C.2 Tavole delle misure di qualità della convergenza

Le tre misure principali — adeguatezza epistemica ε(α), equità η(F), robustezza ϱ(α) — sono definite qui in modo formale e commentate rispetto alla loro applicabilità empirica.

Misura 1. Adeguatezza epistemica

Siano:

- E(N) l'insieme delle proposizioni epistemicamente rilevanti per l'atto α che sono note ad almeno un agente $i \in N$;
- E(Conv) il sottoinsieme di tali proposizioni che ha effettivamente influenzato il processo di convergenza che ha prodotto α.

L'adeguatezza epistemica della convergenza è:

$\varepsilon(\alpha) = |E(\text{Conv}) \cap E(N)| \; / \; |E(N)| \in [0, 1]$

Una convergenza è epistemicamente adeguata nella misura in cui $\varepsilon(\alpha) \rightarrow 1$. Il caso $\varepsilon(\alpha) = 0$ corrisponde a una convergenza completamente cieca alle informazioni disponibili nel sistema; il caso $\varepsilon(\alpha) = 1$ (ideale regolativo) corrisponde alla piena mobilizzazione dell'informazione rilevante.

Misura 2. Equità della convergenza (indice di Gini sui pesi pragmatici)

Siano w_1, w_2, …, w_n ≥ 0 i pesi pragmatici assegnati agli n agenti dalla funzione di convergenza, con Σ(i=1..n) w_i = 1. L'equità della convergenza è:

$$\eta(F) = 1 - \mathrm{Gini}(w_1, \ldots, w_n) = 1 - [\Sigma(i=1..n)\ \Sigma(j=1..n)\ |w_i - w_j|] / [2n \cdot \Sigma(i=1..n)\ w_i] \in [0, 1]$$

Il valore $\eta = 1$ corrisponde alla perfetta equità (tutti i pesi uguali a $1/n$); il valore $\eta = 0$ corrisponde alla dittatura assoluta (un agente ha peso 1, tutti gli altri peso 0).

Misura 3. Robustezza della convergenza

Sia D una distribuzione di probabilità sulle perturbazioni del sistema (modifiche delle preferenze di singoli agenti, ingresso di nuovi agenti, variazioni procedurali). Per ogni perturbazione $\delta \sim D$, sia α_δ l'output che il sistema produrrebbe dopo la perturbazione, e $d(\cdot,\cdot)$ una metrica sullo spazio Σ degli output comunicativi con massimo d_{max}.

La robustezza è:

$$\varrho(\alpha) = 1 - E_{\delta \sim D}[\, d(\alpha, \alpha_\delta) / d_{max} \,] \in [0, 1]$$

Il valore $\varrho = 1$ indica che il sistema produce lo stesso output indipendentemente dalle perturbazioni (convergenza rigida); il valore $\varrho = 0$ indica che qualunque perturbazione infinitesima produce l'output massimalmente distante (convergenza instabile).

Misura 4. Deficit epistemico

Sia α^* l'output che il sistema produrrebbe se la funzione di convergenza fosse epistemicamente ottimale, ossia se $\varepsilon(\alpha^*) = 1$ (tutti i segnali rilevanti mobilizzati con peso proporzionale alla loro qualità epistemica).

$$\delta_{ep}(\alpha) = d(\alpha, \alpha^*) \in [0, d_{max}]$$

Il deficit epistemico misura quanto l'output effettivo si discosta dall'ideale epistemico. Nota che δ_{ep} non è osservabile direttamente (richiederebbe di conoscere α^*), ma è un utile costrutto teorico per valutare ex post i guasti delle strutture di convergenza.

Misura 5. Legittimità procedurale

$$\lambda(\alpha) = 1[\mathrm{Valid_P}(C, \alpha)] = \{\ 1 \text{ se } \mathrm{Valid_P}(C, \alpha) \text{ è vera; } 0 \text{ altrimenti } \}$$

La legittimità procedurale è una misura binaria: l'atto è o non è strutturalmente valido. A differenza delle altre misure, non ammette gradi. Può tuttavia essere estesa a una misura graduata $\lambda_grad(\alpha) \in [0,1]$ pesando il numero di condizioni procedurali soddisfatte sul totale di quelle richieste.

Tavola sinottica delle misure

Misura	**Simbolo**	**Formula compatta**	**Dominio**	**Ottimo**
Adeguatezza epistemica	$\varepsilon(\alpha)$	$\lvert E(Conv) \cap E(N) \rvert \,/\, \lvert E(N) \rvert$	[0, 1]	$\varepsilon = 1$
Equità (anti-Gini)	$\eta(F)$	$1 - Gini(w_1,\ldots,w_n)$	[0, 1]	$\eta = 1$
Robustezza	$\varrho(\alpha)$	$1 - E_\delta[d(\alpha, \alpha_\delta)/d_max]$	[0, 1]	$\varrho = 1$
Deficit epistemico	$\delta_ep(\alpha)$	$d(\alpha, \alpha^*)$	[0, d_max]	$\delta_ep = 0$
Legittimità procedurale	$\lambda(\alpha)$	$1[Valid_P(C, \alpha)]$	{0, 1}	$\lambda = 1$

Nota metodologica. Le misure ε, ϱ e λ sono computazionalmente trattabili in linea di principio ma richiedono dati raramente disponibili nella forma necessaria per i casi reali. La misura δ_ep richiede la definizione di un output di riferimento α^*, non banale in assenza di consenso sullo standard epistemico appropriato. Nonostante queste difficoltà operative, le misure hanno valore normativo e comparativo anche quando non sono calcolabili con precisione numerica.

C.3 Schema delle strutture di convergenza tipiche e loro proprietà attese

Struttura	**ε attesa**	**η attesa**	**ϱ attesa**	**Rischi principali**
Gerarchia pura	Bassa–media	Bassa	Media	groupthink verticale; filtro

				gerarchico dell'informazione
Collegialità paritaria	Media	Alta	Bassa	paradosso del docente; ciclicità condorcetiana
Deliberazione habermasiana (ideale)	Alta	Alta	Alta	irrealistica; costi temporali proibitivi
Procedure Delphi	Alta	Media–alta	Media	lentezza; difficoltà di implementazione istituzionale
Mercato (aggregazione di prezzi)	Media	Bassa	Media	bolle speculative; cascate informative
Algoritmo AI — LLM	Media–bassa	Molto bassa	Alta	distorsioni di corpus; hallucination strutturale
Voto a maggioranza	Media	Alta	Media	paradosso di Condorcet; manipolazione dell'agenda
Consenso (potere di veto)	Bassa	Alta	Alta	inerzia decisionale; sovrarappresentazione delle minoranze di blocco

---

C.4 Il modello DeGroot di aggiornamento delle credenze su reti

In un sistema di n agenti connessi da una rete pesata, il modello DeGroot descrive l'aggiornamento iterativo delle credenze come segue.

Sia $x_i^{(t)} \in \mathbb{R}$ la credenza dell'agente i allo stadio temporale t, e sia $W = (w_{ij})$ la matrice dei pesi, con $w_{ij} \geq 0$ e $\Sigma_{(j=1..n)} w_{ij} = 1$ per ogni i.

Regola di aggiornamento:

$$x_i^{(t+1)} = \Sigma_{(j=1..n)} w_{ij} \cdot x_j^{(t)} \quad \forall\ i \in \{1,\ldots,n\},\ \forall\ t \geq 0$$

In forma matriciale, denotando $\mathbf{x}^{(t)} = (x_1^{(t)}, \ldots, x_n^{(t)})^T$:

$$\mathbf{x}^{(t)} = W^t \cdot \mathbf{x}^{(0)}$$

Teorema di convergenza (DeGroot 1974). Se la matrice W è stocastica per righe (ogni riga somma a 1), irriducibile e aperiodica, allora esiste un vettore di consenso $x^* \in \mathbb{R}$ tale che:

$$\lim_{t\to\infty} x_i^{(t)} = x^* \quad \forall\ i \in \{1,\ldots,n\}$$

dove $x^* = \boldsymbol{\pi}^T \cdot \mathbf{x}^{(0)}$ e $\boldsymbol{\pi}$ è il vettore stazionario sinistro di W (la distribuzione stazionaria della catena di Markov indotta).

Interpretazione pragmatica. Il valore di consenso x^* è una media pesata delle credenze iniziali, con pesi determinati dalla struttura della rete (il vettore $\boldsymbol{\pi}$). Gli agenti con alta centralità — nel senso della misura PageRank, che coincide con π_i per catene di Markov appropriate — esercitano un'influenza sproporzionata sul consenso finale. Strutture di rete altamente centralizzate (es. stelle) producono convergenze in cui un singolo nodo hub determina quasi interamente x^*; strutture uniformi producono convergenze più democratiche.

C.5 Il teorema di Arrow nella pragmatica distribuita: formalizzazione

Sia L lo spazio degli ordini totali stretti su Σ (lo spazio degli output comunicativi possibili). Un profilo di preferenze è una tupla $\mathbf{R} = (R_1, R_2, \ldots, R_n) \in L^n$, dove R_i è l'ordine di preferenza dell'agente i. Una funzione di convergenza è una mappa $F: L^n \to L$ che associa a ogni profilo un ordine collettivo.

Definizione (proprietà desiderabili di F):

- **(U) Dominio universale:** $\mathrm{Dom}(F) = L^n$ F è definita per ogni profilo di preferenze individuali possibile.

- **(P) Principio di Pareto:** $\forall\ \alpha, \beta \in \Sigma, (\forall\ i: \alpha \succ_i \beta) \Rightarrow \alpha \succ_{F(\mathbf{R})} \beta$

- **(IIA) Indipendenza dalle alternative irrilevanti:** $\forall\ \alpha, \beta \in \Sigma, (\forall\ i: R_i{\restriction}\{\alpha,\beta\} = R'_i{\restriction}\{\alpha,\beta\}) \Rightarrow F(\mathbf{R}){\restriction}\{\alpha,\beta\} = F(\mathbf{R}'){\restriction}\{\alpha,\beta\}$ dove ${\restriction}\{\alpha,\beta\}$ denota la restrizione dell'ordine alle sole alternative α e β.

- **(ND) Non-dittatura:** $\neg \exists\ i^* \in \{1,\ldots,n\}$ tale che $\forall\ \mathbf{R} \in L^n$, $\forall\ \alpha, \beta \in \Sigma: \alpha \succ_{\{i^*\}} \beta \Rightarrow \alpha \succ_{F(\mathbf{R})} \beta$

Teorema di Arrow (1951). Per $|\Sigma| \geq 3$ e $n \geq 2$, non esiste nessuna funzione $F: L^n \to L$ che soddisfi simultaneamente le proprietà (U), (P), (IIA) e (ND).

Corollario per la pragmatica normativa. Nessuna struttura di convergenza può essere simultaneamente:

- epistemicamente rispettosa (U): applicabile a qualsiasi configurazione di informazioni;
- rispettosa del consenso (P): quando tutti concordano, il sistema segue;
- invariante rispetto al framing (IIA): immune dalla manipolazione dell'agenda;
- democratica (ND): priva di un agente determinante.

Ogni struttura reale viola almeno una di queste proprietà. La scelta di quale violare non è una scelta tecnica ma una scelta normativa che richiede giustificazione pubblica esplicita nei termini della giustizia comunicativa elaborata nell'Ottava Parte.

C.6 Definizione formale del gioco di convergenza e del nucleo

Definizione (gioco di convergenza). Un gioco di convergenza è una quadrupla:

$G = (N, \Sigma, \{u_i\}_{i \in N}, P)$

dove:

- $N = \{1, 2, \ldots, n\}$ è l'insieme finito degli agenti partecipanti;
- Σ è lo spazio degli output comunicativi possibili (l'insieme di tutti gli atti linguistici candidati);
- $u_i : \Sigma \to \mathbb{R}$ è la funzione di utilità dell'agente i, che assegna a ogni output un valore di preferenza;
- P è l'insieme delle procedure ammissibili — le regole che il processo deve seguire per produrre un output strutturalmente valido.

Definizione (funzione caratteristica del gioco di convergenza). Per ogni coalizione $S \subseteq N$, il valore $v(S)$ è la massima utilità aggregata che la coalizione S può garantire ai propri membri attraverso le procedure P_S ammissibili per S:

$$v(S) = \max_{\sigma \in \Sigma,\, P \in P_S} \Sigma_{i \in S}\, u_i(\sigma)$$

Definizione (nucleo del gioco di convergenza). Il nucleo C(G) è l'insieme degli output $\sigma^* \in \Sigma$ tali che nessuna coalizione propria può bloccare σ^*, nel senso che:

$$C(G) = \{\, \sigma^* \in \Sigma \mid \neg \exists\, S \subsetneq N,\, \sigma' \in \Sigma : \forall\, i \in S,\, u_i(\sigma') > u_i(\sigma^*) \wedge \sigma' \text{ è } P_S\text{-ammissibile} \,\}$$

Un output nel nucleo è stabile rispetto alle deviazioni di coalizioni: nessun gruppo di agenti ha la capacità procedurale e la motivazione a divergere dall'accordo collettivo.

Teorema (Bondareva–Shapley, adattato). Il nucleo C(G) è non vuoto se e solo se il gioco (N, v) è bilanciato, ossia per ogni sistema bilanciato di coefficienti $\{\lambda_S\}_{\{S \subseteq N\}}$ *con* $\Sigma_{\{S \ni i\}}\, \lambda_S = 1 \;\forall\; i \in N$ e $\lambda_S \geq 0$, vale:

$$\Sigma_{S \subseteq N}\, \lambda_S \cdot v(S) \leq v(N)$$

Interpretazione. Un gioco bilanciato è uno in cui nessuna combinazione di coalizioni parziali può ottenere aggregativamente di più di quanto otterrebbe la grande coalizione N. La bilanciabilità è la condizione strutturale che garantisce la possibilità di una convergenza stabile: sistemi in cui questa condizione non è soddisfatta producono strutture di convergenza intrisecamente instabili, in cui qualunque output può essere minacciato da deviazioni di sotto-gruppi.

Appendice D

Glossario critico della pragmatica distribuita

Atto linguistico distribuito. Atto linguistico prodotto da un sistema di agenti multipli attraverso un processo di convergenza, senza che esista un centro intenzionale unitario che lo generi. Proprietà: ha forza illocutoria determinata (spesso procedurale), condizioni di felicità strutturali e sostanziali potenzialmente dissociate, responsabilità distribuita tra i partecipanti al processo di convergenza. Si distingue dagli atti linguistici collettivi per i quali si assume una intenzione condivisa: l'atto linguistico distribuito non presuppone né richiede tale condivisione. Formalmente: Conv(N, α) con $\neg \exists\, i \in N : I_i\, \alpha \wedge \forall\, j \neq i : \neg I_j\, \alpha$.

Adeguatezza epistemica (di una convergenza). Misura $\varepsilon(\alpha) \in [0,1]$ della proporzione di informazioni rilevanti disponibili nel sistema distribuito che sono state effettivamente mobilizzate dal processo di convergenza. Una convergenza è epistemicamente adeguata nella misura in cui $\varepsilon(\alpha)$ si avvicina a 1. Si distingue dalla validità strutturale (conformità procedurale, $\lambda(\alpha) \in \{0,1\}$) con cui può dissociarsi: $\lambda(\alpha) = 1$ è compatibile con $\varepsilon(\alpha) \approx 0$.

Ambiguità intenzionale distribuita. Fenomeno per cui un atto linguistico prodotto da un sistema distribuito contiene ambiguità strutturale — non riducibile a vaghezza semantica o contestuale — che riflette un conflitto non risolto tra sotto-sistemi in competizione nel processo di convergenza. L'ambiguità intenzionale distribuita è spesso un meccanismo deliberato di convergenza: consente a sotto-sistemi con preferenze incompatibili di accordarsi su un testo comune rinunciando alla determinatezza del contenuto. Formalizzabile come: $\exists\ S_1, S_2 \subsetneq N : u_{S_1}(\alpha_1) > u_{S_1}(\alpha_2)$ e $u_{S_2}(\alpha_2) > u_{S_2}(\alpha_1)$, con α che "comprende" entrambe le interpretazioni α_1 e α_2.

Assunzione di stabilità funzionale. Sostituto, nella pragmatica del distribuito, dell'assunzione di cooperatività griceana. Il destinatario di un atto linguistico distribuito può razionalmente assumere che il sistema che lo ha prodotto abbia un interesse funzionale nel mantenere la coerenza tra i propri atti linguistici e le proprie pratiche, non per benevolenza verso il destinatario ma per ragioni di autoriproduzione istituzionale.

Cascata informativa. Fenomeno in cui gli agenti di un sistema distribuito razionalmente ignorano le proprie informazioni private s_i e

si conformano al comportamento della maggioranza, producendo una convergenza collettiva che può essere sistematicamente errata. Si formalizza come situazione in cui il segnale aggregato $\Sigma_{j<i} 1[a_j = a^*]$ domina il segnale privato s_i nel calcolo bayesiano di ogni agente i della sequenza. La cascata produce alta robustezza ϱ ma bassa adeguatezza epistemica ε: robustezza e qualità epistemica si disaccoppiano.

Condizioni di felicità strutturali. Condizioni la cui soddisfazione determina la validità procedurale $\lambda(\alpha) = 1$ di un atto linguistico distribuito: conformità alle procedure P, rispetto dei requisiti di quorum e competenza, osservanza delle forme previste. Si distinguono dalle condizioni di felicità sostanziali, con cui possono dissociarsi.

Condizioni di felicità sostanziali. Condizioni che riguardano la fondatezza epistemica del processo di convergenza: che i contribuenti avessero le credenze e le informazioni necessarie per produrre un atto del tipo in questione. Formalmente: SubstFel(α) è vera quando $\varepsilon(\alpha)$ supera una soglia normativa contestualmente determinata. Corrispondono alla condizione di sincerità di Searle, riformulata in termini di accesso informativo del sistema piuttosto che di stati mentali del parlante.

Convergenza (funzione di). Processo $F: \{v_1, \ldots, v_n\} \to O$ che trasforma un insieme di vettori di preferenza degli agenti in un output comunicativo O dotato di forza illocutoria determinata, contenuto proposizionale determinato, e condizioni di soddisfazione determinate. La funzione di convergenza è caratterizzata dalla sua struttura (gerarchica, procedurale, negoziale, algoritmica), dalle sue proprietà (ε, η, ϱ), e dall'algoritmo attraverso cui è realizzata.

Deficit epistemico. Distanza $\delta_{ep}(\alpha) = d(\alpha, \alpha^*)$ tra l'output della convergenza effettiva α e l'output epistemicamente ottimale α^* (quello che il sistema produrrebbe con $\varepsilon = 1$). Misura quanto la struttura di convergenza si discosta dall'ideale della verità di sistema.

Dipendenza dal sentiero (della convergenza). Proprietà di molte strutture di convergenza per cui l'output finale è sensibile all'ordine temporale in cui i contributi vengono integrati. Formalizzabile attraverso l'estensione dinamica del gioco di convergenza con stadio $t \in T$: $\sigma_T \neq \sigma'_T$ anche quando $\{v_1, \ldots, v_n\}$ è identico ma l'ordine di integrazione

differisce. La dipendenza dal sentiero genera path-locking e rende il processo manipolabile attraverso il controllo dell'agenda.

Distribuzione agentiva. Forma di distribuzione in cui più agenti $i_1, \ldots, i_k \in N$ concorrono alla produzione di α senza che nessuno lo produca interamente: $\mathrm{Conv}(\{i_1,\ldots,i_k\}, \alpha)$ con $\neg \exists\, j : \mathrm{Conv}(\{j\}, \alpha)$. Si distingue dalla distribuzione deliberativa (il vettore v_i di ogni agente è esso stesso il prodotto di un processo di convergenza interno) e dalla distribuzione cognitiva (le risorse K_i sono distribuite e nessun agente possiede K_N).

Equità della convergenza. Misura $\eta(F) = 1 - \mathrm{Gini}(w_1,\ldots,w_n) \in [0,1]$ della distribuzione dei pesi pragmatici w_i tra gli n agenti. $\eta = 1$ indica pesi uguali ($w_i = 1/n$ per tutti); $\eta = 0$ indica dittatura assoluta.

Forza illocutoria procedurale. Forza illocutoria assegnata dalla struttura istituzionale all'output del processo di convergenza $F(\mathbf{R})$, indipendentemente dalle intenzioni psicologiche dei partecipanti: $\mathrm{Valid_P}(N, \alpha) \rightarrow \mathrm{Force}(\alpha) = f_P$, dove f_P è la forza illocutoria associata alla procedura P. Una sentenza ha forza di cosa giudicata anche se il giudice non ha compreso appieno le implicazioni della propria decisione.

Funzione di convergenza. Vedi: Convergenza (funzione di).

Gioco di convergenza. Modello formale del processo di produzione di un atto linguistico distribuito, definito come la quadrupla $G = (N, \Sigma, \{u_i\}_{i\in N}, P)$ (si veda la Sezione C.6 per la definizione completa).

Groupthink. Patologia della convergenza in cui pressioni alla conformità all'interno di un gruppo coeso producono $\varepsilon(\alpha) \ll 1$ pur mantenendo $\varrho(\alpha) \approx 1$: il sistema è stabile ma epistemicamente cieco. Formalizzabile come situazione in cui i pesi w_{ij} della matrice DeGroot sono fortemente concentrati all'interno dei cluster, producendo una convergenza intra-cluster rapida e una convergenza inter-cluster assente.

Implicatura distribuita. Inferenza che il destinatario trae applicando l'assunzione di stabilità funzionale: dato che il sistema ha interesse a mantenere coerenza tra atti e pratiche, l'asserzione $\mathrm{Assert}(N, \varphi)$ implica contestualmente $K_N\, \varphi$, ossia che il sistema dispone di informazioni che giustificano φ. L'implicatura si incrina quando l'interesse funzionale del sistema si disallinea dall'interesse epistemico.

Ingiustizia epistemica distribuita. Violazione del principio di proporzionalità del peso: $w_i \ll \varepsilon_i$, dove ε_i è il contributo epistemico genuino dell'agente i al pool informativo del sistema. Il sistema struttura la convergenza assegnando peso pragmatico inferiore al peso epistemico a certe categorie di agenti, per ragioni di status e non di qualità informativa.

Interprete dell'emisfero sinistro. Meccanismo sub-personale (Gazzaniga) che produce narrazioni causali post-hoc coerenti a partire da comportamenti generati da processi inaccessibili alla coscienza. Modello micro-livello della funzione di convergenza intra-personale: il soggetto individuale è esso stesso un sistema distribuito che produce la propria rappresentazione intenzionale come output $F(v_1,\ldots,v_k)$ di processi sub-personali $v_1, \ldots, v_k$ non accessibili alla coscienza.

Male strutturale comunicativo. Produzione sistematica di atti linguistici epistemicamente degradati o eticamente problematici attraverso strutture di convergenza che nessun singolo agente ha intenzionalmente progettato per produrre quel tipo di output: formalmente, $\delta_{ep}(\alpha) > \theta$ per qualche soglia normativa θ, con $\neg \exists\, i \in N : I_i\, \mathrm{DegradedOutput}(\alpha)$.

Nucleo (del gioco di convergenza). L'insieme $C(G)$ degli output stabili (si veda la definizione formale in C.6). La non-vacuità del nucleo è condizione sufficiente per la robustezza $\varrho(\alpha) > 0$ della convergenza.

Policromia illocutoria. Proprietà $\exists\, S_1, S_2 \subsetneq N : \mathrm{Force}_{S_1}(\alpha) \neq \mathrm{Force}_{S_2}(\alpha)$: l'atto α ha forze illocutorie distinte per sotto-sistemi diversi che vi hanno contribuito. Caso speciale di ambiguità intenzionale distribuita in cui la divergenza riguarda la forza illocutoria piuttosto che il contenuto proposizionale.

Principio del soggetto unitario (PSU). Assunzione che ogni atto linguistico sia riconducibile a un agente singolare con struttura psicologica sufficientemente integrata da consentire l'attribuzione di intenzioni $I_i\, \alpha$ discrete, coscienti e causalmente efficaci. Formalmente: PSU implica $\exists!\, i : \mathrm{Author}(\alpha, i)$. Presupposto implicito della pragmatica classica che la pragmatica distribuita rimpiazza con la funzione di convergenza.

Responsabilità epistemica distribuita. Secondo livello del framework: $\mathrm{Resp}(i, \alpha)$ proporzionalmente a $K_i\, \mathrm{contrib}(i, \alpha)$ e a quanto i poteva

sapere delle implicazioni dell'atto finale. Si veda l'assioma (A4) della DMAL.

Responsabilità per design della convergenza. Terzo livello: RespDesign(i, P), la responsabilità di chi progetta le procedure P per gli output sistematici che queste strutture producono. Distinta dalla responsabilità per singoli atti: riguarda le caratteristiche statistiche degli output dell'intera classe di convergenze prodotte da P.

Responsabilità strutturale. Primo livello: RespStruct(N, α) dell'ente per l'atto α prodotto da una convergenza sotto il suo controllo normativo. Si veda l'assioma (A6) della DMAL: Valid_P(N, α) ∧ Control(N, P) → RespStruct(N, α).

Robustezza (della convergenza). Misura ϱ(α) ∈ [0,1] della stabilità dell'output rispetto a perturbazioni δ ~ D. Si veda la definizione formale in C.2. Alta robustezza corrisponde a convergenze nel nucleo C(G) o in prossimità di esso.

Semiotica post-intenzionale. Teoria del segno che rinuncia all'intenzione del soggetto come condizione necessaria del significato. Le quattro tesi fondamentali: (i) il significato è posizionale — determinato dalla posizione nella struttura semiotica (Sign, Object, Interpretant) di Peirce — indipendentemente dall'intenzione del parlante; (ii) è processuale — il risultato di un processo di convergenza interpretativa distribuita; (iii) è normativo — governato da pratiche condivise che vincola l'uso corretto; (iv) è emergente — può eccedere le intenzioni di tutti i partecipanti al processo.

Significato emergente. Significato m(α) di un atto linguistico distribuito tale che m(α) ∉ ∪_{i∈N} {m_i(α)}, dove m_i(α) è il significato inteso dall'agente i. Il significato emergente è analizzabile attraverso la struttura di F ma non è riducibile alla somma delle intenzioni individuali.

Stabilità funzionale (assunzione di). Vedi: Assunzione di stabilità funzionale.

Sotto-sistema in competizione. Componente S ⊊ N del sistema distribuito con funzione di utilità u_S che può divergere da u_{N∖S}. La teoria dei sotto-sistemi in competizione modella il processo di convergenza come un gioco tra sotto-sistemi il cui esito F(**v**_S) è determinato dalla struttura di potere relativa (pesi pragmatici w_S) e dalle procedure ammissibili P.

Testimonianza distribuita. Atto Assert(N, φ) prodotto da un sistema distribuito N con funzione di credibilità Cred(N, φ) dipendente dalla qualità della struttura di convergenza: Cred(N, φ) = f(ε(φ), η(F)), con f crescente in entrambi gli argomenti. Un sistema con alta adeguatezza epistemica e alta equità produce testimonianza distribuita più credibile.

Verità di sistema. Proposizione φ tale che φ sarebbe asserita dall'output α* della convergenza epistemicamente ottimale (ε = 1) del sistema N. Formalmente: φ è vera-di-sistema per N se $\vDash_{E(N)}$ φ, ossia se φ è implicata dalla totalità delle informazioni disponibili nel sistema. La distanza tra verità di sistema e output effettivo è il deficit epistemico $\delta_{ep}(\alpha) = d(\alpha, \alpha^*)$.

Bibliografia

Opere citate nel testo e riferimenti fondamentali

Austin, J.L. (1962). *How to Do Things with Words*. Oxford University Press. [Edizione italiana: *Come fare cose con le parole*, Marietti, 1987. La raccolta delle William James Lectures del 1955, fondativa della teoria degli atti linguistici. La distinzione tra constativo e performativo e la teoria delle condizioni di felicità rimangono punti di partenza imprescindibili, analizzati criticamente nella Prima Parte.]

Arrow, K.J. (1951). *Social Choice and Individual Values*. Wiley. [Il teorema di impossibilità è formalizzato nella Prima edizione. La seconda edizione del 1963 contiene chiarimenti importanti. Rilevante per la Sesta Parte come fondamento dell'impossibilità di una funzione di convergenza che soddisfi simultaneamente tutte le proprietà desiderabili.]

Aumann, R.J. (1976). Agreeing to disagree. *Annals of Statistics*, 4(6), 1236–1239. [Teorema fondamentale sulla conoscenza comune: agenti con probabilità comuni non possono "concordare di discordare". Rilevante per la modellazione formale della convergenza nella Sesta Parte.]

Bikhchandani, S., Hirshleifer, D., Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, 100(5), 992–1026. [Fondativa della teoria delle cascate informative, discussa nella Settima Parte come patologia della convergenza epistemica.]

Butler, J. (1990). *Gender Trouble: Feminism and the Subversion of Identity*. Routledge. [La teoria della performatività di genere, anticipazione della teoria della performatività in *Bodies That Matter* (1993). Il contributo al problema della performatività istituzionale è discusso nella Quarta Parte.]

Butler, J. (1993). *Bodies That Matter: On the Discursive Limits of "Sex"*. Routledge. [Sviluppo sistematico della teoria della performatività. Discusso nella Quarta Parte come tradizione complementare alla teoria degli atti linguistici.]

Butler, J. (1997). *Excitable Speech: A Politics of the Performative*. Routledge. [Analisi della violenza del discorso e della responsabilità per atti linguistici che producono danni. Rilevante per la Quarta Parte sulla responsabilità distribuita.]

Clark, A., Chalmers, D. (1998). The extended mind. *Analysis*, 58(1), 7–19. [Proposta della teoria della mente estesa e del principio di parità. Discusso nell'Appendice A come base teorica per la cognizione estesa nelle organizzazioni.]

Clark, A. (2016). *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford University Press. [Sviluppo sistematico della teoria del predictive processing. Rilevante per l'Appendice A sulla teoria cognitiva della produzione linguistica.]

Coady, C.A.J. (1992). *Testimony: A Philosophical Study*. Oxford University Press. [Opera fondativa dell'epistemologia della testimonianza. Discusso nella Settima Parte come punto di partenza per la nozione di testimonianza distribuita.]

Cohen, J. (1989). Deliberation and democratic legitimacy. In A. Hamlin, P. Pettit (eds.), *The Good Polity*. Blackwell, 17–34. [Articolazione della teoria deliberativista della democrazia. Discusso nella Quarta Parte in relazione alle conseguenze della pragmatica distribuita per la filosofia politica.]

Cyert, R.M., March, J.G. (1963). *A Behavioral Theory of the Firm*. Prentice Hall. [Fondativa dell'analisi delle organizzazioni come coalizioni con preferenze divergenti. Centrale per la Terza Parte sulla distribuzione deliberativa.]

DeGroot, M.H. (1974). Reaching a consensus. *Journal of the American Statistical Association*, 69(345), 118–121. [Il modello di aggiornamento bayesiano su reti che porta il nome dell'autore è discusso nella Sesta Parte come algoritmo di convergenza.]

Dworkin, R. (1977). *Taking Rights Seriously*. Harvard University Press. [Il modello di Hercules come giudice ideale è rilevante come caso limite dell'integrazione della funzione di convergenza distribuita in una volontà unitaria fittizia. Discusso nella Terza e Quarta Parte.]

Dworkin, R. (1986). *Law's Empire*. Harvard University Press. [La teoria dell'integrità del diritto come risposta coerente alla complessità della produzione giuridica distribuita è un punto di confronto importante per la Quarta Parte.]

Elster, J. (1998). (ed.) *Deliberative Democracy*. Cambridge University Press. [Raccolta che articola il programma della democrazia deliberativa in relazione alle pratiche istituzionali reali. Rilevante per la Quarta Parte.]

Financial Crisis Inquiry Commission (2011). *The Financial Crisis Inquiry Report*. U.S. Government Printing Office. [Documentazione fondamentale per l'analisi del caso delle agenzie di rating nella Quinta Parte.]

Foucault, M. (1966). *Les mots et les choses*. Gallimard. [L'analisi dell'episteme come sistema di regole che determina cosa può essere detto e pensato in un'epoca è rilevante per la Sesta e Nona Parte sulla semiotica del distribuito.]

Foucault, M. (1969). *L'archéologie du savoir*. Gallimard. [La teoria del discorso come pratica che forma gli oggetti di cui parla è un riferimento per la semiotica post-intenzionale della Nona Parte.]

Foucault, M. (1971). *L'ordre du discours*. Gallimard. [La lezione inaugurale al Collège de France sull'ordine del discorso e le procedure di controllo. Citato nel corpo principale.]

Fricker, M. (2007). *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford University Press. [Fondativa della nozione di ingiustizia epistemica. L'estensione al contesto dei sistemi distribuiti è elaborata nella Settima e Ottava Parte.]

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127–138. [Esposizione sintetica del framework del predictive processing. Rilevante per l'Appendice A.]

Gazzaniga, M.S. (1985). *The Social Brain: Discovering the Networks of the Mind*. Basic Books. [Opera principale sulla teoria dell'interprete dell'emisfero sinistro. Citata nel corpo principale e approfondita nell'Appendice A.]

Gazzaniga, M.S. (2011). *Who's in Charge? Free Will and the Science of the Brain*. Ecco Press. [Sintesi accessibile delle implicazioni delle ricerche sul cervello diviso per la teoria dell'intenzione e del controllo. Discusso nell'Appendice A.]

Goldman, A. (1999). *Knowledge in a Social World*. Oxford University Press. [Sviluppo sistematico dell'epistemologia sociale. Rilevante per la Settima Parte.]

Gramsci, A. (1975). *Quaderni del carcere*. Einaudi. [L'analisi dell'egemonia come controllo del discorso pubblico è un riferimento per la Quarta Parte sulla critica ideologica del linguaggio istituzionale.]

Grice, H.P. (1957). Meaning. *Philosophical Review*, 66(3), 377–388. [Prima formulazione della teoria del significato-nn. Citato nella Prima Parte.]

Grice, H.P. (1975). Logic and conversation. In P. Cole, J. Morgan (eds.), *Syntax and Semantics, vol. 3: Speech Acts*. Academic Press, 41–58. [La formulazione canonica del Principio di Cooperazione e delle massime conversazionali. Citato nel corpo principale.]

Habermas, J. (1981). *Theorie des kommunikativen Handelns*. Suhrkamp. [La teoria dell'azione comunicativa e l'ideale del discorso. Citato nel corpo principale. Discusso criticamente nella Prima e Quarta Parte.]

Habermas, J. (1992). *Faktizität und Geltung*. Suhrkamp. [La democrazia deliberativa come risposta alla crisi di legittimità dello stato di diritto. Rilevante per la Quarta Parte.]

Halliday, M.A.K. (1978). *Language as Social Semiotic*. Edward Arnold. [La teoria delle funzioni del linguaggio (ideazionale, interpersonale, testuale) è discussa nella Nona Parte come base per la sociosemiotica del distribuito.]

Halliday, M.A.K., Hasan, R. (1976). *Cohesion in English*. Longman. [La teoria della coesione testuale è rilevante per l'analisi dell'intertestualità distribuita nella Nona Parte.]

Hart, H.L.A. (1961). *The Concept of Law*. Oxford University Press. [La distinzione tra regole primarie e secondarie e la teoria del punto di vista interno sono riferimenti per la Quarta Parte sulla filosofia del diritto.]

Haynes, J.D., et al. (2007). Reading hidden intentions in the human brain. *Current Biology*, 17(4), 323–328. [Studio che estende i risultati di Libet mostrando capacità predittive del comportamento fino a 10 secondi prima della consapevolezza. Discusso nell'Appendice A.]

Janis, I. (1972). *Victims of Groupthink*. Houghton Mifflin. [Opera fondativa del concetto di groupthink come patologia della convergenza nelle decisioni di gruppo. Discusso nella Settima Parte.]

Kornhauser, L., Sager, L. (1986). Unpacking the court. *Yale Law Journal*, 96(1), 82–117. [Prima formulazione del problema del judgment aggregation nel contesto giuridico, poi formalizzato da List e Pettit come paradosso del docente. Discusso nella Settima Parte.]

Kress, G., van Leeuwen, T. (1996). *Reading Images: The Grammar of Visual Design*. Routledge. [La semiotica multimodale come estensione dell'approccio sociosemiotico. Rilevante per la Nona Parte.]

Lackey, J. (2008). *Learning from Words: Testimony as a Source of Knowledge*. Oxford University Press. [Contributo importante all'epistemologia della testimonianza. Rilevante per la Settima Parte sulla testimonianza distribuita.]

Libet, B., Gleason, C.A., Wright, E.W., Pearl, D.K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). *Brain*, 106(3), 623–642. [Gli esperimenti fondativi

sul Readiness Potential. Citati nel corpo principale e discussi nell'Appendice A.]

List, C., Pettit, P. (2002). Aggregating sets of judgments: an impossibility result. *Economics and Philosophy*, 18(1), 89–110. [Prima formalizzazione del paradosso del docente (discursive dilemma) come problema di aggregazione dei giudizi. Discusso nella Settima Parte.]

List, C., Pettit, P. (2011). *Group Agency: The Possibility, Design, and Status of Corporate Agents*. Oxford University Press. [Trattazione sistematica della questione dell'agency collettiva. Centrale per la Terza e Quarta Parte sulla responsabilità distribuita.]

Lotman, J.M. (1984). *O semiosfere*. Trudy po znakovym sistemam / Sign Systems Studies, 17, 5–23. [La nozione di semiosfera come spazio semiotico strutturato è discussa nella Nona Parte.]

Lotman, J.M. (1990). *Universe of the Mind: A Semiotic Theory of Culture*. Tauris. [Raccolta degli scritti fondamentali sulla semiosfera e sulla dinamica dei sistemi semiotici. Citato nella Nona Parte.]

March, J.G., Olsen, J.P. (1976). *Ambiguity and Choice in Organizations*. Universitetsforlaget. [Il modello garbage can delle decisioni organizzative è rilevante per la Terza Parte sulla distribuzione deliberativa.]

March, J.G., Simon, H.A. (1958). *Organizations*. Wiley. [Opera fondativa della teoria comportamentale dell'organizzazione. Discusso nella Terza Parte.]

Mouffe, C. (1993). *The Return of the Political*. Verso. [La critica del deliberativismo razionalista in nome del carattere agonistico del politico è rilevante per la Quarta Parte.]

Olson, M. (1965). *The Logic of Collective Action*. Harvard University Press. [Il problema del free-rider nella produzione di beni pubblici è rilevante per la comprensione degli incentivi che strutturano i processi di convergenza.]

Peirce, C.S. (1931–1958). *Collected Papers*, voll. 1–8. Harvard University Press. [La semiotica triadica e la distinzione icona/indice/simbolo sono discusse nella Nona Parte. I volumi più rilevanti sono il 2 (Elements of Logic) e il 5 (Pragmatism and Pragmaticism).]

Pettit, P. (2001). *A Theory of Freedom: From the Psychology to the Politics of Agency*. Oxford University Press. [La teoria della libertà come non-interferenza è rilevante per l'analisi della responsabilità degli agenti distribuiti.]

Rawls, J. (1971). *A Theory of Justice*. Harvard University Press. [Il velo di ignoranza come dispositivo di convergenza epistemica è un riferimento per la discussione dei principi di giustizia comunicativa nella Ottava Parte.]

Rawls, J. (1993). *Political Liberalism*. Columbia University Press. [La ragione pubblica e le condizioni di legittimità democratica. Citato nel corpo principale. Discusso nella Quarta Parte.]

Scanlon, T.M. (1998). *What We Owe to Each Other*. Harvard University Press. [Il contrattualismo di Scanlon è rilevante per i principi etici del design delle strutture di convergenza nella Ottava Parte.]

Searle, J.R. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press. [La sistemazione della teoria degli atti linguistici di Austin con la distinzione tra regole costitutive e regolative. Citato nel corpo principale.]

Searle, J.R. (1983). *Intentionality: An Essay in the Philosophy of Mind*. Cambridge University Press. [La psicologizzazione sistematica degli atti linguistici come forme di intenzionalità. Rilevante per la Prima Parte.]

Searle, J.R. (1995). *The Construction of Social Reality*. Free Press. [La teoria delle status functions e dell'intenzionalità collettiva. Citata nel corpo principale e discussa criticamente nella Prima e Terza Parte.]

Searle, J.R. (2010). *Making the Social World: The Structure of Human Civilization*. Oxford University Press. [Sviluppo e difesa della teoria della realtà sociale istituzionale. Citato nel corpo principale.]

Shapley, L.S., Shubik, M. (1954). A method for evaluating the distribution of power in a committee system. *American Political Science Review*, 48(3), 787–792. [L'indice di Shapley-Shubik come misura del potere nelle strutture di voto è rilevante per la formalizzazione della funzione di convergenza nella Sesta Parte.]

Simon, H.A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, 69(1), 99–118. [La nozione di razionalità limitata (bounded rationality). Citato nel corpo principale.]

Simon, H.A. (1957). *Administrative Behavior*. Macmillan. [Il fondamento teorico dell'analisi del processo decisionale nelle organizzazioni. Rilevante per la Terza Parte.]

Soon, C.S., Brass, M., Heinze, H.J., Haynes, J.D. (2008). Unconscious determinants of free decisions in the human brain. *Nature Neuroscience*, 11(5), 543–545. [Studio che mostra predittività dell'attività cerebrale per decisioni binarie fino a 10 secondi. Discusso nell'Appendice A.]

Sperry, R.W. (1961). Cerebral organization and behavior. *Science*, 133(3466), 1749–1757. [Gli studi fondativi sul cervello diviso. Discussi nell'Appendice A.]

Sunstein, C.R. (2009). *Going to Extremes: How Like Minds Unite and Divide*. Oxford University Press. [L'analisi della polarizzazione di gruppo è rilevante per la Settima Parte sui meccanismi di bolla epistemica nelle strutture di convergenza.]

Sunstein, C.R., Hastie, R. (2015). *Wiser: Getting Beyond Groupthink to Make Groups Smarter*. Harvard Business Review Press. [Analisi empirica e normativa dei meccanismi di decisione collettiva. Rilevante per la Settima e Ottava Parte.]

Surrowiecki, J. (2004). *The Wisdom of Crowds*. Doubleday. [La documentazione empirica del fenomeno della saggezza delle folle è rilevante per la Settima Parte sulla conoscenza emergente nei sistemi distribuiti.]

van Dijk, T.A. (1998). *Ideology: A Multidisciplinary Approach*. Sage. [L'analisi critica del discorso come strumento di critica ideologica. Citato nel corpo principale.]

van Dijk, T.A. (2008). *Discourse and Power*. Palgrave Macmillan. [Sviluppo della teoria dell'analisi critica del discorso con attenzione esplicita alle relazioni di potere. Rilevante per la Quarta e Ottava Parte.]

Weick, K.E. (1979). *The Social Psychology of Organizing*. Random House. [La teoria dell'enactment e del sensemaking nelle organizzazioni è rilevante per la Terza Parte.]

Weick, K.E. (1995). *Sensemaking in Organizations*. Sage. [Il concetto di sensemaking come processo di costruzione retrospettiva del significato nelle organizzazioni è parallelo alla nozione di interprete di Gazzaniga e rilevante per la Nona Parte sul significato emergente.]

Wittgenstein, L. (1953). *Philosophische Untersuchungen*. Blackwell. [La critica al linguaggio privato e la nozione di Lebensform come fondamento del significato nella pratica collettiva. Citato nel corpo principale e rilevante per la Nona Parte.]

Wittgenstein, L. (1969). *On Certainty*. Blackwell. [La struttura del dubbio e della certezza come pratiche sociali. Rilevante per la Settima Parte sull'epistemologia del distribuito.]

Young, I.M. (2006). Responsibility and global justice: a social connection model. *Social Philosophy and Policy*, 23(1), 102–130. [La distinzione tra responsabilità individuale e responsabilità politica per l'ingiustizia strutturale. Citata nella Ottava Parte.]

Young, I.M. (2011). *Responsibility for Justice*. Oxford University Press. [Sviluppo sistematico del modello di responsabilità sociale per le strutture. Centrale per la concezione di responsabilità distribuita nella Ottava Parte.]

Zagzebski, L. (2012). *Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief*. Oxford University Press. [La teoria dell'autorità epistemica è rilevante per la comprensione del ruolo dei sistemi distribuiti come fonti di conoscenza testimoniale.]

Riferimenti normativi e documentali

Commissione Europea (2022). *Proposta di Regolamento del Parlamento Europeo e del Consiglio sull'intelligenza artificiale (AI Act)*. COM(2021) 206 final, aggiornato dalla posizione del Parlamento Europeo del marzo 2024. [Discusso nella Quinta Parte come caso di atto linguistico distribuito con ambiguità strutturale deliberata.]

Commissione Europea (2022). *Proposta di Direttiva del Parlamento Europeo e del Consiglio sulla responsabilità nell'intelligenza artificiale (AI Liability Directive)*. COM(2022) 496 final. [Discussa nell'Appendice B come sviluppo del diritto della responsabilità degli enti per output di sistemi AI.]

Decreto Legislativo 8 giugno 2001, n. 231. Disciplina della responsabilità amministrativa delle persone giuridiche, delle società e delle associazioni anche prive di personalità giuridica. [Discusso nell'Appendice B come sistema di responsabilità degli enti.]

Financial Crisis Inquiry Commission (2011). *The Financial Crisis Inquiry Report: Final Report of the National Commission on the Causes of the Financial and Economic Crisis in the United States*. U.S. Government Printing Office. [Documentazione primaria per il caso delle agenzie di rating nella Quinta Parte.]

IPCC (2021). *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press. [Discusso nella Quinta Parte come caso di produzione scientifica collettiva e struttura di convergenza distribuita.]

Letteratura recente e studi emergenti

Acemoglu, D., Robinson, J.A. (2012). *Why Nations Fail: The Origins of Power, Prosperity and Poverty*. Crown. [L'analisi delle istituzioni estrattive come strutture che producono sistematicamente certi tipi di output comunicativi è rilevante per la Quarta Parte.]

Barocas, S., Hardt, M., Narayanan, A. (2019). *Fairness and Machine Learning*. fairmlbook.org. [L'analisi del fairness nei sistemi di machine learning è rilevante per la valutazione delle strutture di convergenza degli algoritmi AI nella Sesta Parte.]

Bender, E.M., Gebru, T., McMillan-Major, A., Shmitchell, S. (2021). On the dangers of stochastic parrots: can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (FAccT). [Analisi critica dei rischi dei large language models. Rilevante per la Seconda e Quinta Parte.]

Binns, R. (2018). Fairness in machine learning: lessons from political philosophy. *Proceedings of Machine Learning Research*, 81, 1–11. [Il collegamento tra teoria politica della giustizia e fairness algoritmica è rilevante per la Settima e Ottava Parte.]

Braithwaite, J., Drahos, P. (2000). *Global Business Regulation*. Cambridge University Press. [L'analisi dei sistemi di regolazione globale come strutture di convergenza distribuita a livello internazionale è rilevante per la Quarta e Quinta Parte.]

Dafoe, A. (2018). AI governance: a research agenda. Future of Humanity Institute, University of Oxford. [L'agenda di ricerca sulla governance dell'AI è un riferimento per le questioni aperte della Decima Parte.]

Floridi, L., et al. (2018). AI4People — an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4), 689–707. [Il framework etico per l'AI è un riferimento per la Ottava Parte sull'etica delle strutture di convergenza computazionali.]

Frith, C. (2007). *Making Up the Mind: How the Brain Creates Our Mental World*. Blackwell. [La teoria della mente come costruzione cerebrale è rilevante per l'Appendice A sulla distribuzione sub-personale della cognizione.]

Goldberg, S. (2010). *Relying on Others: An Essay in Epistemology*. Oxford University Press. [L'epistemologia della reliance on others è rilevante per la Settima Parte sulla testimonianza distribuita.]

Hagendorff, T. (2020). The ethics of AI ethics: an evaluation of guidelines. *Minds and Machines*, 30(1), 99–120. [Analisi critica dei framework etici per l'AI. Rilevante per la Ottava Parte.]

Jackson, M., Rogers, B.W., Zenou, Y. (2017). The economic consequences of social network structure. *Journal of Economic Literature*, 55(1), 49–95. [Survey della letteratura sulle reti sociali con attenzione agli effetti economici. Rilevante per la Sesta Parte sulla modellazione a reti della funzione di convergenza.]

Jobin, A., Ienca, M., Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399. [Analisi

comparativa dei principi etici per l'AI nelle principali iniziative internazionali. Rilevante per la Ottava Parte.]

Kleinberg, J., Ludwig, J., Mullainathan, S., Sunstein, C.R. (2018). Human decisions and machine predictions. *Quarterly Journal of Economics*, 133(1), 237–293. [L'analisi comparativa delle decisioni umane e algoritmiche è rilevante per la Sesta Parte sugli algoritmi di convergenza.]

Krakovna, V., et al. (2020). Specification gaming: the flip side of AI ingenuity. *DeepMind Blog*. [I fenomeni di specification gaming nei sistemi AI come manifestazioni di convergenza algoritmica disallineata dai valori intesi sono rilevanti per la Seconda e Ottava Parte.]

Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S., Floridi, L. (2016). The ethics of algorithms: mapping the debate. *Big Data and Society*, 3(2). [Mappa del dibattito etico sugli algoritmi. Rilevante per la Ottava Parte.]

Noble, S.U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York University Press. [L'analisi delle distorsioni razziali nei sistemi algoritmici come forme di ingiustizia epistemica distribuita è rilevante per la Settima e Ottava Parte.]

O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown. [L'analisi critica degli algoritmi decisionali come strutture di convergenza con distorsioni sistematiche è rilevante per la Ottava Parte.]

Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press. [La critica della opacità degli algoritmi decisionali è rilevante per il principio di trasparenza della struttura nella Ottava Parte.]

Rahwan, I. (2018). Society-in-the-loop: programming the algorithmic social contract. *Ethics and Information Technology*, 20(1), 5–14. [Il framework "society in the loop" per la governance dei sistemi AI è rilevante per la Decima Parte sul programma di ricerca.]

Russell, S. (2019). *Human Compatible: Artificial Intelligence and the Problem of Control*. Viking. [La proposta di allineamento dell'AI agli obiettivi umani come problema di design della funzione di convergenza è rilevante per la Seconda e Ottava Parte.]

Sloman, S., Fernbach, P. (2017). *The Knowledge Illusion: Why We Never Think Alone*. Riverhead Books. [L'analisi della cognizione come intrinsecamente distribuita anche a livello individuale è rilevante per l'Appendice A.]

Wachter, S., Mittelstadt, B., Russell, C. (2017). Counterfactual explanations without opening the black box: automated decisions and the GDPR. *Harvard Journal of Law and Technology*, 31(2), 841–887. [Il problema della spiegabilità delle decisioni algoritmiche è rilevante per il principio di tracciabilità nella Settima Parte.]

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs. [L'analisi del capitalismo della sorveglianza come sistema che produce atti linguistici — previsioni comportamentali — attraverso strutture di convergenza computazionali opache è rilevante per la Quinta e Ottava Parte.]

Nota sulla struttura bibliografica

La presente bibliografia è organizzata in quattro sezioni: opere citate nel testo e riferimenti fondamentali, riferimenti normativi e documentali, letteratura recente e studi emergenti. Le annotazioni a ciascuna voce indicano la parte o le parti della monografia in cui il riferimento è utilizzato e la natura del suo contributo teorico.

La sezione "letteratura recente e studi emergenti" include contributi pubblicati dopo il 2000 che non erano disponibili agli autori della tradizione classica analizzata nella Prima Parte, ma che sono rilevanti per le questioni aperte identificate nella Decima Parte e per lo sviluppo del programma di ricerca della pragmatica distribuita. Questa sezione non è esaustiva: la letteratura rilevante — particolarmente nel settore dell'etica dell'AI e della governance algoritmica — si sviluppa a un ritmo che rende impossibile qualsiasi aggiornamento definitivo.

Un segnale di questa velocità di sviluppo: al momento di chiusura del presente lavoro, la letteratura sull'AI Act europeo e sulla AI Liability Directive è ancora in formazione, e i riferimenti normativi citati potrebbero avere ricevuto aggiornamenti significativi nel processo legislativo. La pragmatica distribuita, come scienza dell'architettura comunicativa, è per sua natura un campo in cui la teoria deve inseguire

una realtà che si trasforma più rapidamente di quanto le monografie accademiche possano documentare.